



सांख्यिकी-विमर्श

2017-18

अंक
13

संस्थान द्वारा संयुक्त रूप से विकसित पोर्टल ऐप



भा.क.अनु.प. भारतीय कृषि सांख्यिकी अनुसंधान संस्थान
लाइब्रेरी एबेन्यू, पुरा, नई दिल्ली 110 012
www.iasri.res.in
आईएसओ 9001:2008 प्रमाणित संस्थान
आईएसओ/आईसी 20000 तथा 27001 प्रमाणित डाटा केंद्र



सांख्यिकी-विमर्श

2017-18

अंक
13

संपादक मण्डल		
	अजीत	अध्यक्ष
1	अनिल कुमार	सदस्य
2	ऊषा जैन	
3	द्विजेश चन्द्र मिश्र	
4	देवी प्रसाद शर्मा	
5	ब्रह्मजीत गहलौत	
6	सुशील कुमार सरकार	
7	संतोष कुमार	
8	संतोष कुमार सिंह	

आवरण एवं अन्तिम पृष्ठ अवधारणा एवं डिजाइनिंग
अजीत



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

लाइब्रेरी एवेन्यू, पूसा, नई दिल्ली - 110012

www.iasri.res.in

आईएसओ 9001: 2008 प्रमाणित संस्थान

आईएसओ/आइईसी 20000 तथा 27001 प्रमाणित डाटा केंद्र



सांख्यिकी-विमर्श

2017-18



संस्थान की वेब साइट का हिन्दी प्रारूप
<http://www.iasri.res.in>

The screenshot shows the homepage of the Indian Agricultural Statistics Research Institute (IASRI). At the top, it features the IASRI logo and the text 'भारतीय कृषि सांख्यिकी अनुसंधान संस्थान'. Below this, there is a navigation menu with categories like 'संस्थान', 'संस्कृत', 'संशोधन', 'संस्थापन एवं प्रशिक्षण', 'विज्ञान योजन प्रोजेक्ट्स', 'परामर्श सेवाएं', 'वार्षिक रिपोर्ट', 'समाचार', 'आर्वाइव्ड', and 'ई-जर्नल्स'. A central image shows a group of people in front of a building, with the caption 'Click the image to take a Virtual Tour'. To the right, there are sections for 'आज के कार्यक्रम', 'विज्ञान 2050', 'समाचार पत्र', 'मा.क.सां.अ.सं. ने आई.सी.टी. की परत', 'केंद्रीय कृषि एवं विज्ञान कल्याण नदी द्वारा के.वी.के. पोर्टल का शुभारंभ', 'सांख्यिकी-विमर्श', 'वर्ष 2014 के वार्षिक संस्करण में अनुसंधान हिन्दी संस्करण की रिपोर्ट', and 'वर्ष 2014 के वार्षिक संस्करण में अनुसंधान हिन्दी संस्करण की रिपोर्ट'. A green box at the bottom of the screenshot contains the text: 'कार्य संवर्धी सभी सम्मेलनों पर विशेष की सलाह के लिये प्रातः 6.00 बजे से रात 10.00 बजे तक सभी सातों दिन टोल फ्री नम्बर 1551 पर संचालन करें ।'.

द्वारा प्रकाशित

निदेशक

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

मुद्रण : मार्च, 2018

भारतीय कृषि सांख्यिकी अनुसंधान संस्थान
INDIAN AGRICULTURAL STATISTICS
RESEARCH INSTITUTE

रूपरेखा एवं मुद्रण : मैसर्स रॉयल ऑफसेट प्रिन्टर्स, ए 89/1 नारायणा इण्डस्ट्रियल एरिया फेस-1,
नई दिल्ली-110028 मो. 9811622258

सम्पादकीय

सांख्यिकीय सभ्यता की गति में अंकों का योगदान बड़ा ही महत्वपूर्ण रहा है और अंक पद्धति के विकास का बहुत बड़ा श्रेय भारत को प्राप्त है। मनुष्य के ज्ञान की प्रत्येक शाखा अंकों का ऋणी है। सांख्यिकी विज्ञान भी बहुत कुछ काम अंकों से लेता है, जिन्हें आँकड़े कहते हैं, परन्तु इन अंकों के विशिष्ट लक्षण होते हैं। व्यष्टि का अध्ययन न करके, समष्टि नाम से अभिहित समूह या समुदाय का अध्ययन करना सांख्यिकी विज्ञान की मौलिक धारणा है। जैसा कि आप सबको विदित है कि सांख्यिकी एक गणितीय विज्ञान है जिसमें किसी वस्तु/अवयव/तंत्र/समुदाय से सम्बन्धित आँकड़ों का संग्रह, विश्लेषण, व्याख्या या स्पष्टीकरण एवम् प्रस्तुती की जाती है। सांख्यिकी विज्ञान एक ऐसी विधा है, जो विज्ञान के हर क्षेत्र चाहे वो चिकित्सा विज्ञान, सामाजिक विज्ञान, व्यवसायिक विज्ञान, प्रबंधन विज्ञान, मनोविज्ञान अथवा कृषि विज्ञान हो, सभी क्षेत्रों में अग्रणी भूमिका निभाती है।

उल्लेखनीय है कि पत्रिका सांख्यिकी-विमर्श के अब तक बारह अंक प्रकाशित हो चुके हैं और अब तेरहवाँ अंक आपके सम्मुख प्रस्तुत किया जा रहा है। सांख्यिकी-विमर्श के इस निरंतर सफलता के पीछे पाठकवृन्द एवम् समस्त लेखकों का योगदान एवम् कौशल निहित है। यह अत्यन्त ही हर्ष का विषय है कि सांख्यिकी-विमर्श लोकप्रिय हो रहा है एवम् इसका लेखों को सराहा जा रहा है। किन्तु हमारे लिए कंवल इतना ही पर्याप्त नहीं है क्योंकि हम जानते हैं कि सांख्यिकी-विमर्श को शीर्षतम उँचाइयों तक ले जाया जा सकता है। इस दिशा में हमारे प्रयास निरंतर जारी हैं और हम उम्मीद करते हैं कि आने वाले दिनों में यह पत्रिका और भी रुचिकर एवम् ज्ञानवर्धक सिद्ध होगी।

सांख्यिकी-विमर्श के इस अंक में सांख्यिकी साहित्य के विभिन्न आयामों यथा सांख्यिकीय जैवमिति, परीक्षण अभिकल्पना, प्रतिदर्श सर्वेक्षण, जैवसूचना एवम् संगणक अनुप्रयोग के लेख प्रस्तुत हैं। हमें उम्मीद ही नहीं अपितु पूर्ण विश्वास है कि सांख्यिकी विमर्श का यह अंक पाठकों के लिए अत्यन्त ज्ञानवर्धक एवम् रोचक सिद्ध होगा।

यह पत्रिका दो खण्डों में प्रकाशित हो रही है, प्रथम खण्ड अनुसंधान खण्ड है जिसमें विज्ञान की जटिलताओं को सुग्राह्य भाषा में लेखों के माध्यम से समझाया गया है। पत्रिका का दूसरा खण्ड राजभाषा खण्ड है, जिसमें साहित्यिक लेख जैसे कि संस्थान में प्रगति के बढ़ते चरण, संस्थान की राजभाषा यात्रा, दैनिक स्मरणीय शब्द शतक, लैटिन शब्दों के हिन्दी समानक एवम् द्विभाषी पदनाम, अन्य स्तम्भ जैसे अतीत के झरोखे से आदि हैं। इस प्रकार हम कह सकते हैं कि यह पत्रिका विज्ञान एवम् साहित्य का अनुपम संगम है, जो इस को एक मौलिक एवम् अतुलनीय पत्रिका बनाता है।

संपादक मंडल "सांख्यिकी-विमर्श" के सभी लेखकों के योगदान के लिए हृदय से आभारी है एवं उम्मीद करता है कि भविष्य में भी उनका अमूल्य सहयोग जारी रहेगा।

इस पत्रिका को समृद्ध एवम् रुचिकर बनाने के लिए, पाठकों के सुझाव अपेक्षित हैं। सभी पाठकों एवम् लेखकों से आग्रह है कि कृपया अपनी प्रतिक्रिया हमें अवश्य भेजें ताकि इस पत्रिका को और अधिक प्रभावशाली ज्ञानवर्धक एवम् रोचक बनाया जा सके। इस पत्रिका के निरंतर विकास की कामना के साथ।

संपादक मंडल



निदेशक की कलम से

वर्ष 2005-06 में संस्थान में हिन्दी पत्रिका 'सांख्यिकी-विमर्श' के प्रकाशन का शुभारम्भ हुआ और तब से अब तक प्रतिवर्ष इस पत्रिका का प्रकाशन किया जा रहा है। प्रस्तुत अंक इस पत्रिका का तेरहवाँ अंक है। संस्थान की हिन्दी पत्रिका "सांख्यिकी-विमर्श 2017-18" आपके समक्ष प्रस्तुत करते हुए मुझे हार्दिक प्रसन्नता की अनुभूति हो रही है।

पत्रिका के इस अंक में संस्थान द्वारा किये गये अनुसंधानों व अन्य कार्यों के संक्षिप्त विवरण, संस्थान में राजभाषा से सम्बन्धित कार्यों आदि की जानकारी के साथ-साथ कृषि सांख्यिकी, कृषि में संगणक अनुप्रयोग एवं कृषि जैव सूचना से सम्बन्धित विभिन्न लेखों एवं शोध-पत्रों को सम्मिलित किया गया है। पत्रिका में पाठकों के हिन्दी ज्ञानवर्धन एवं दैनिक उपयोग के लिए द्विभाषी पदनाम, लैटिन शब्दों के हिन्दी समानक, दैनिक स्मरणीय शब्द-शतक (सांख्यिकी एवं तकनीकी) हिन्दी व अँग्रेजी में दिये गये हैं।

वर्ष 2017-18 की महत्वपूर्ण उपलब्धियों में से प्रमुख इस प्रकार हैं: संस्थान द्वारा विकसित कृषि-शिक्षा-पोर्टल-Agriculture Education Portal (<http://education.icar.gov.in>), पशु प्रजनन-मोबाइल-ऐप (Pashu Prajanan App) तथा सूकर-पालन-मोबाइल-ऐप (Pig Farming App) का शुभारम्भ माननीय श्री राधा मोहन सिंह, केन्द्रीय कृषि एवं कृषक कल्याण मंत्री, कृषि एवं कृषक कल्याण मंत्रालय, भारत सरकार द्वारा 08 फरवरी, 2018 को किया गया। इन दोनों मोबाइल ऐप्स का विकास संस्थान द्वारा आई.वी.आर.आई., इज्जतनगर, बरेली (भारतीय पशु-चिकित्सा अनुसंधान संस्थान, इज्जतनगर) के सहयोग से किया गया है। किसान भाइयों को तकनीकी जानकारी सुगमता से उपलब्ध कराने हेतु संस्थान द्वारा इस वर्ष के.वि.के. (कृषि विज्ञान केन्द्र) मोबाइल-ऐप में अनेक अतिरिक्त सुविधाएँ शामिल की गयी हैं। इन मोबाइल ऐप्स के द्वारा किसान भाइयों को जानकारी उपलब्ध कराने के लिए इनमें अनेक भारतीय भाषाओं का प्रावधान रखा गया है। इसके अतिरिक्त, संस्थान द्वारा इस वर्ष अनेक सांख्यिकीय तकनीकों तथा बायोइन्फॉर्मेटिक्स सर्वर, डाटाबेस तथा पोर्टल का भी विकास किया गया। ये सभी डाटाबेस, सर्वर, पोर्टल संस्थान की वेबसाइट (www.iasri.res.in) पर उपलब्ध हैं।

मैं पत्रिका के प्रकाशन के लिए उन सभी लेखकों का आभारी हूँ जिन्होंने इस पत्रिका में प्रकाशन हेतु अपने लेख देकर हमारे इस प्रयास को सफल बनाने में हमारा सहयोग किया। पत्रिका के प्रकाशन के लिए सम्पादक मंडल के अध्यक्ष, डॉ. अजीत एवं सदस्यों, डॉ. अनिल कुमार, डॉ. द्विजेश चन्द्र मिश्र, डॉ. सुशील कुमार सरकार, सुश्री ऊषा जैन, श्री संतोष कुमार, श्री संतोष कुमार सिंह, श्री ब्रह्मजीत गहलौत तथा श्री देवी प्रसाद शर्मा का मैं आभार व्यक्त करता हूँ जिनके अथक प्रयासों से यह पत्रिका इस रूप में आपके समक्ष आ सकी।

आशा है इस अंक की विषय-वस्तु पाठकों के लिए सूचनाप्रद एवं उपयोगी सिद्ध होगी और सांख्यिकी जैसे तकनीकी विषय में भी हिन्दी साहित्य का प्रयोग करके पाठकों का ज्ञानवर्धन करने में सहयोगी सिद्ध होगी। इस पत्रिका के भावी अंकों में सुधार के लिए आपके सुझावों का स्वागत है।

लालमोहन
(लाल मोहन भर)
निदेशक

सांख्यिकी-विमर्श

2017-18

अंक
13

राजभाषा सम्मान



नराकास (उत्तरी दिल्ली) द्वारा आयोजित प्रश्नोत्तरी प्रतियोगिता के प्रथम चरण में संस्थान के प्रतिभागियों द्वारा द्वितीय स्थान प्राप्त करने पर संस्थान के निदेशक प्रमाण-पत्र प्राप्त करते हुए



अनुक्रमणिका

सम्पादकीय	iii
निदेशक की कलम से	v
अनुसंधान खण्ड	
संस्थान में शोध प्रगति के बढ़ते चरण	1
भारत में कश्मीर के पर्वतीय कुपवाड़ा जनपद में प्रचलित कृषिवानिकी पद्धतियों के द्वारा संचित कार्बन स्टॉक तथा कार्बन अवशोषण क्षमता का आंकलन	9
<ul style="list-style-type: none"> • अजीत, अनिल कुमार, ब्रह्मजीत गहलौत, विजय पाल सिंह, नरेश कुमार एवं उमा बकरियों में वृद्धि वक्र प्राचलों की आनुवंशिक परिवर्तनशीलता : बूट-स्ट्रेप तकनीकों का अनुप्रयोग 	15
<ul style="list-style-type: none"> • ए.के. पॉल, रंजित कुमार पॉल, लाल मोहन भर, हिमाद्रि शेखर राय, प्रकाश कुमार, सविता वधवा एवं सत्यपाल सिंह 	
कृषि में अरफिमा मॉडल और उसके अनुप्रयोग	24
<ul style="list-style-type: none"> • दिपांकर मित्र, रंजित कुमार पॉल, लाल मोहन भर, ए. के. पॉल, हिमाद्रि शेखर राय, सविता वधवा एवं ए. आर. उदगाता 	
विविधता के दो स्रोतों के अंतर्गत पशु परीक्षणों के लिए प्रवृत्ति मुक्त अभिकल्पनाएँ	28
<ul style="list-style-type: none"> • अर्पण भौमिक, सीमा जग्गी, एल्दो वरगीस, सुनील कुमार यादव, मोहम्मद हारुन, अनिदिता दत्ता एवं उदयवीर सिंह 	
माइक्रोसैटेलाइट डी.एन.ए. मार्कर के प्रयोग से मवेशी नस्ल पहचान हेतु वेब सर्वर	32
<ul style="list-style-type: none"> • मीर आसिफ इकबाल, सारिका, ऊषा जैन, अनिल राय एवं दिनेश कुमार 	
कृषि में कीट संख्या की गतिशीलता पर आई.एन.ए.आर. (INAR) मॉडल का अनुप्रयोग	37
<ul style="list-style-type: none"> • हिमाद्रि शेखर राय, रंजित कुमार पॉल, एल. एम. भर, मृन्मय राय, प्रकाश कुमार, अचल लामा, दीपंकर मित्र एवं राजु कुमार 	
एरिमा काल श्रृंखला मॉडलिंग	45
<ul style="list-style-type: none"> • रामसुब्रमणियन वी., रविन्द्र सिंह शेखावत, के. एन. सिंह एवं एस. पी. सिंह 	
ए. एन. एन. का अवलोकन: उत्तर प्रदेश के झांसी जिले की मानसून वर्षा की भविष्यवाणी में अनुप्रयोग	50
<ul style="list-style-type: none"> • मृन्मय राय, के. एन. सिंह, प्रकाश कुमार, राजीव रंजन कुमार, संतोष राठोड़, प्रदीप बसाक, हिमाद्रि शेखर राय एवं अचल लामा 	
एस ए एस द्वारा आंकड़ों की सारांश रिपोर्ट	55
<ul style="list-style-type: none"> • राजेंद्र सिंह तोमर एवं प्रमोद कुमार 	
सी 4.5 एल्गोरिथ्म का प्रयोग करते हुए निर्णय-वृक्ष वर्गीकरण एवं विजुअलाइजेशन के लिए एक ऑनलाइन सॉफ्टवेयर (ओडीटीसी)	65
<ul style="list-style-type: none"> • शशि दहिया, सुवजीत दास, अंशु भारद्वाज एवं ऊषा जैन 	
सपोर्ट वेक्टर समाश्रयण से काल श्रृंखला आंकड़ों का पूर्वानुमान: भारतीय तिलहन उत्पादन डेटा में अनुप्रयोग	71
<ul style="list-style-type: none"> • संतोष राठोड़, के. एन. सिंह, प्रकाश कुमार, राजीव रंजन कुमार, सुशील कुमार सरकार, विशाल गुरुंग, मृन्मय राय एवं अचल लामा 	

अनुपात का लघु क्षेत्र आकलन	75
• सुशील कुमार, हुकुम चन्द्र, वन्दिता कुमारी, कौस्तव आदित्य एवं प्रदीप बसाक	
सर्वेक्षण डाटा विश्लेषण में सर्वेक्षण भार का महत्व	81
• वन्दिता कुमारी, हुकुम चन्द्र, सुशील कुमार, प्रदीप बसाक, अंकुर विश्वास एवं कौत्सव आदित्य	
रैपिड माइनर के द्वारा टेक्स्ट माइनिंग	87
• कृष्ण कुमार चतुर्वेदी, अनिल राय, मो. समीर फारुकी, शशि भूषण लाल, अनु शर्मा, यू. बी. अंगडी एवं विशाल मिश्र	
फज्जी तकनीक से पशुधन निर्यात का विश्लेषण	94
• रविन्द्र सिंह शेखावत, के. एन. सिंह, विशाल गुरुंग राजीव रंजन कुमार, अचल लामा, प्रवीण आर्य एवं मो. यासीन	
विषम चरों के लिए समष्टि योग की प्रागुक्ति	96
• प्रदीप बसाक, हुकुम चन्द्र, कौस्तव आदित्य, मृन्मय राय, वन्दिता कुमारी, सुशील कुमार एवं ऊषा जैन	
स्वतंत्रता प्राप्ति के बाद भारत में चावल के उत्पादन की स्थिति एवं पूर्वानुमान	101
• वसी आलम, रंजीत के. पॉल, प्रवीण आर्या एवं ऊषा जैन	
विकसित प्रतिमान का आकलन करने के लिए प्रसंभाव्य विभिन्नात्मक समीकरण पद्धति	104
• हिमाद्री घोष, सविता कधवा एवं प्रज्ञेषु	
स्थिर लाइनों के उत्पादन के लिए उच्च स्तरीय क्रॉसिंग योजनाएँ	109
• मोहम्मद हारून, सिनी वरगीस, सीमा जग्गी, अनिदिता दत्ता एवं अर्पण भौमिक	
मेटाजीनोमिक्स : परिचय और अनुप्रयोग	113
• अनु शर्मा, शशि भूषण लाल, द्विजेश चन्द्र मिश्र, नीरज बुढ़लाकोटी, कृष्ण कुमार चतुर्वेदी, मोहम्मद समीर फारुकी एवं संजीव कुमार	
भारतीय कृषि अनुसंधान परिषद (आईसीएआर) में एंटरप्राइज रिसोर्स प्लानिंग सिस्टम (आईसीएआर-ईआरपी) का कार्यान्वयन	116
• मुकेश कुमार, सुदीप मारवाह, अलका अरोड़ा, अंशु भारद्वाज, अंजनी कुमार चौबे एवं शशि दहिया	

राजभाषा खण्ड

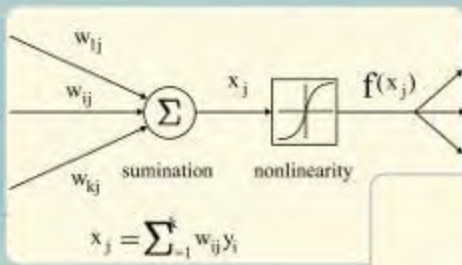
संस्थान की राजभाषा यात्रा : 2017-18	127
द्विभाषी पदनाम	131
लैटिन शब्दों के हिन्दी समानक	133
दैनिक स्मरणीय शब्द-शतक	134
अतीत के झरोखे से...	136
हिन्दी पखवाड़ा-2017 तथा अन्य हिन्दी कार्यक्रम	141
वर्ष 2017-18 के दौरान संस्थान में हिन्दी कार्यशालाओं का आयोजन	148
ऑर्गेनोग्राम	154

सांख्यिकी-विमर्श

2017-18

अंक
13

अनुसंधान खण्ड



संस्थान में शोध प्रगति के बढ़ते चरण

भा.कृ.अनु.प.-भारतीय कृषि सांख्यिकी अनुसंधान संस्थान (भा.कृ.अनु.प.-भा.कृ.सां.अ.सं) मौजूदा ज्ञान में विकास एवं वृद्धि के लिए कृषि सांख्यिकी एवं सूचना विज्ञान में अनुसंधान कार्य के लिए प्रयासरत है। संस्थान ने सांख्यिकी का प्रयोग विज्ञान के रूप में किया है तथा इसके साथ सूचना विज्ञान का प्रासंगिक प्रयोग किया है और कृषि अनुसंधान की गुणवत्ता को और अधिक उन्नत बनाने में महत्वपूर्ण योगदान दिया है। संस्थान द्वारा देश में प्रशिक्षित मानवशक्ति प्रदान करने के लिए कृषि सांख्यिकी और सूचना विज्ञान के क्षेत्र में शिक्षा/प्रशिक्षण भी प्रदान किया जाता है। कृषि अनुसंधान में गुणवत्ता सुधारने तथा उभरते हुए नए क्षेत्रों की चुनौतियों से निपटने के लिए सांख्यिकी अनुसंधान एवं शिक्षा का प्रयोग किया जाता है। प्रतिवेदनाधीन वर्ष में संस्थान द्वारा विकसित किये गये कृषि शिक्षा पोर्टल, प्रजनन मोबाइल ऐप तथा सूकर पालन मोबाइल ऐप प्रमुख हैं। इसके अतिरिक्त कृषि-पोर्टल तथा के.वी.के. पोर्टल एवं के.वी.के. मोबाइल ऐप में अनेक अतिरिक्त सुविधाओं को भी शामिल किया गया है। भा.कृ.सां.अ.सं में नेटवर्किंग सेवाओं को और अधिक सुदृढ़ किया गया। संस्थान का सम्पूर्ण परिसर हाई स्पीड इंटरनेट कनेक्शन के साथ वाई-फाई सक्षम है, ताकि संस्थान के छात्र और कर्मचारी किसी भी स्थान से इंटरनेट का प्रयोग कर सकें। भा.कृ.अनु.प. में एकीकृत मेल मैसेजिंग एवं वेब होस्टिंग समाधान उपलब्ध कराने हेतु संस्थान में भा.कृ.अनु.प. डाटा सेंटर (टियर-11) स्थापित किया गया है।

संस्थान ने कृषि सांख्यिकी एवं सूचना विज्ञान के अनुसंधान में उत्कृष्ट एवं महत्वपूर्ण योगदान दिया है। प्रतिवेदनाधीन वर्ष के दौरान संस्थान ने अनेक अनुसंधान परियोजनाएँ चलाई। संस्थान में कुल 92 अनुसंधान परियोजनाओं के अंतर्गत अनुसंधान कार्य किया गया (इनमें से 19 संस्थान द्वारा वित्त-पोषित,

18 अन्य बाह्य एजेन्सियों द्वारा वित्त-पोषित, 01 राष्ट्रीय अध्येता स्कीम तथा नयी परियोजनाय 29 थी (जिसमें 11 संस्थान पोषित तथा 18 बाह्य वित्त पोषित) एवं 25 परियोजनायें पूर्ण हुई।

संस्थान की कुछ विशेष अनुसंधानिक उपलब्धियाँ निम्न हैं:

- प्रायः यह वांछनीय होता है कि परीक्षणों को अनुक्रिया पृष्ठ में इस प्रकार निष्पादित किया जाए कि इनपुट कारक स्तर परिवर्तनों की संख्या को लघु रखा जाए ताकि उपकरण की लागत को किफायती बनाया जा सके। प्लाक्केट-बर्मन अभिकल्पनाओं के रन अनुक्रमों में परिवर्तनों की कुल संख्या के लिए एक सामान्य व्यंजकता प्राप्त की गई और एसएसएस मैक्रो का प्रयोग करते हुए प्लाक्केट-बर्मन अभिकल्पनाएँ सृजित की गईं। अनुक्रिया पृष्ठ अभिकल्पनाओं (आरएसडी) के लिए न्यूनतम परिवर्तित रन अनुक्रमों के सृजन के लिए एक आर-सॉफ्टवेयर पैकेज विकसित किया गया, नामतः *मिनिमल आरएसडी* और उसे <https://cran.r-project.org/web/packages/minimalRSD/index.html> पर होस्ट किया गया। इस पैकेज में अभिकल्पनाओं की तीन श्रृंखलाएँ हैं, अर्थात् पूर्ण एवं भिन्नात्मक बहुउपादानी प्वाइंटों (हाफ रेप्लीकेट) और न्यूनतम परिवर्तित रन अनुक्रमों के साथ बॉक्स बेहनकेन अभिकल्पनाएँ (आरबीडी)।
- भा.कृ.अनु.प. के विभिन्न संस्थानों में उपलब्ध/विकसित अनेक ऑनलाइन संसाधनों के लिंक उपलब्ध कराकर KRISHI पोर्टल (<http://krishi.icar.gov.in>) को समृद्ध किया गया। 15 प्रमुख शीर्षकों के तहत कृषि पोर्टल पर भा.कृ.अनु.प. और अन्य संसाधनों के लिंकों की कुल संख्या 269 है। विभिन्न संगठनों द्वारा भा.कृ.अनु.प. के अंतर्गत और बाहर से एकत्रित अनेक कृषि शोध प्रकाशनों के



लिए इंटरपोर्टल हार्वेस्टर (<http://krishi.icar.gov.in/ohs-2.3.1/>) को सुदृढ़ किया गया। मेटाडाटा हार्वेस्टिंग (ओएआई-पीएमएच) प्रोटोकाल समर्थित वेब अनुप्रयोगों, अर्थात् कृषि कोश, सीएमएफआरआई ई-प्रिंट्स, एनएएआरएम प्रिंट्स, आईआईएसआर में DSpace, भारतीय कृषि अनुसंधान जर्नल, OAR@ICRISAT, आईसीआरआईएसएटी डाटावर्स, आईआईटी, मुंबई में DSpace: एसआईआर-एनआईएससीआईआर; सीएसआईआर-मद्रास कॉम्पलेक्स; एनआईएससीआईआर ऑनलाइन पिरियॉडिकल रिपोजिटरी; शोधगंगा (छात्रों को अपने पीएच. डी. शोध प्रबंध प्रस्तुत करने के लिए एक प्लेटफॉर्म), सीआईएफटी में DSpace; सीएफटीआरआई में ई-प्रिंट्स; भाकृअनुप KRISHI प्रकाशन एवं डाटा इन्वेंटरी रिपोजिटरी, भारतीय विज्ञान संस्थान का इलेक्ट्रॉनिक शोध प्रबंध; भारतीय शोध प्रबंध की ओपन ऐक्सेस रिपोजिटरी; सीएसआईआर इत्यादि के लिए 17 ओपन आर्काइव्स इनिशियेटिव प्रोटोकॉल से मेटा डाटा हार्वेस्ट किया गया। 1,84,466 अभिलेखों हेतु 17 रिपोजिटरियों के लिए एकीकृत सर्च तैयार है। उपयोगकर्ताओं को समझने में सहजता प्रदान करने के लिए सभी स्थानिक लेयर्स को सिंगल व्यू में समाविष्ट कर भाकृअनुप-जियोपोर्टल वर्जन विकसित किया गया और उपयोगकर्ताओं के प्रश्नों को प्रदर्शित करने हेतु स्थानिक लेयर "प्रश्न" के साथ इसे कार्यान्वित किया गया। कृषि पोर्टल पर 125 देशों (गुगल विश्लेषण के अनुसार) के 1500 शहरों से 38,000 से अधिक लोगों ने संपर्क किया है।

- कंट्रोल ट्रीटमेंट्स के एक सेट के साथ टेस्ट ट्रीटमेंट्स के एक सेट की तुलना करने हेतु A-इष्टतम संतुलित बाइपारटाइट (BBPB) तथा भारत A-इष्टतम ब्लॉक अभिकल्पनाओं की संरचना के लिए आर.लैंग्वेज का प्रयोग करते हुए एक एल्गोरिथ्म (कलनविधि) विकसित किया गया और प्राचलीकृत रेंज $v_1 < 30, v_2 < 3, b < 50, k < 10$ जहाँ v_1 टेस्ट ट्रीटमेंट्स की संख्या को, v_2 कंट्रोल की संख्या को, b ब्लॉकों की संख्या को तथा k ब्लॉक

आकार की संख्या को दर्शाता है, में A-इष्टतम बीबीपीबी अभिकल्पनाओं एवं भारत A-इष्टतम ब्लॉक अभिकल्पनाओं की एक सूची की संरचना के लिए एल्गोरिथ्म का प्रयोग किया गया।

- संगणक अनुकारकों पर संचालित परीक्षणों तथा वैश्विक संवेदनशील विश्लेषण के लिए लेटिन हाइपरक्यूब अभिकल्पनाएँ लोकप्रिय विकल्प सिद्ध हुईं। किसी भी अनुज्ञेय रन्स के साथ 2, 3, 4, 5 और 6 कारकों के साथ ऑर्थोगोनल लेटिन हाइपरक्यूब अभिकल्पनाओं के निर्माण हेतु संपूर्ण समाधान उपलब्ध कराने के लिए संरचना निर्माण विधियाँ विकसित की गईं।
- कृषि प्रणाली में शिक्षण एवं ई-लर्निंग प्रक्रिया की महत्ता तथा कृषि शिक्षा में आईसीटी टूल्स पर इसके प्रभाव की परिकल्पना को ध्यान में रखते हुए, एक वेब आधारित अनुप्रयोग तथा एक वेबसाइट विकसित की गई, नामतः ICTIAA (<http://ictiaa.icar.gov.in>)। यह वेबसाइट MOOC (व्यापक ओपन ऑनलाइन पाठ्यक्रम) तथा आईसीटी टूल्स का प्रभाव विश्लेषण, दोनों सेवाएँ उपलब्ध कराती है।
- पूर्ण मुख्य प्लॉटों और अपूर्ण उप-प्लॉटों के साथ अपूर्ण स्प्लिट प्लॉट अभिकल्पनाएँ प्राप्त करने के लिए एक संरचना विधि विकसित की गई।
- एक अध्ययन के अंतर्गत आनुभविक रूप से यह समीक्षा करने के लिए कि क्या कृषि जिनसों (एग्रिकल्चरल कमोडिटीज़) की ट्रेडिंग में बाजार दक्षता एवं पारदर्शिता हासिल करने हेतु बेहतर मूल्य प्राप्त करने के लिए वायदा बाजार सहायता करते हैं या नहीं, यह निष्कर्ष निकाला गया कि प्राइस डिस्कवरी केवल वायदा बाजार से ही प्राप्त की जा सकती है, न कि स्पॉट बाजार से। दूसरे शब्दों में, हम यह कह सकते हैं कि मूल्य स्थापन में स्पॉट बाजार के मूल्य के बजाय, वायदा बाजार की भूमिका अधिक होती है। इस अध्ययन में यह दर्शाया गया है कि इष्टतम हैज अनुपात का संगणन करने में समय परिवर्ती प्राचल मॉडल, अचर प्राचल हैज अनुपात मॉडल की तुलना में,



बेहतर है। इस वास्तविकता की पुष्टि दो हैज अनुपात मॉडल्स के ग्राफिकल प्रस्तुतीकरण में की गई है। अध्ययन में यह भी दर्शाया गया है कि समय परिवर्ती मॉडल में संगणित इष्टतम है, वायदा अभिपत्रों में बड़े निवेश में सहायता करता है।

- स्टार्मा मॉडल को कर्नाटक के नौ जिलों में मौसमीय समायोजित मासिक अधिकतम तापमान में कार्यान्वित किया गया। औसत निरपेक्ष प्रतिशत त्रुटि (एमएपीई) के आधार पर बहुचर ऐरिमा एवं स्टार्मा मॉडल के परिणामों की तुलना की गई और यह पाया गया कि बहुचर ऐरिमा मॉडल की तुलना में, स्टार्मा मॉडल का निष्पादन प्रत्येक स्थान में बेहतर था। स्टार्मा मॉडल के प्राचलों के इष्टतमीकरण के लिए मशीन लर्निंग इष्टतमीकरण तकनीकों, जैसे कि जेनेटिक एल्गोरिथ्म (जीए) और पार्टिकल स्वार्म ऑप्टिमाइजेशन (पीएसओ) का प्रयोग किया गया। अध्ययन में यह पाया गया कि वर्तमान स्टार्मा मॉडल की तुलना में, प्रस्तावित पद्धति का निष्पादन बेहतर था।
- वीएआर मॉडल में पूर्वानुमान अंतराल की संरचना के लिए सीव बूटस्ट्रेप पद्धति विकसित की गई। प्रस्तावित पद्धति को यथार्थ एवं अनुकारित डाटासेट्स में कार्यान्वित किया गया। आनुभविक परिणामों में यह दर्शाया गया है कि किंफिडेंस इंटरवल की लेंथ के आधार पर पारंपरिक पद्धति की तुलना में प्रस्तावित पद्धति श्रेष्ठ थी। न्यूरल तंत्रिका में पूर्वानुमान अंतराल की संरचना के लिए सीव बूटस्ट्रेप पद्धति विकसित की गई। प्रस्तावित पद्धति को मैके ग्लास समीकरण से सृजित अनुकारित डाटासेट्स में कार्यान्वित किया गया।
- जीओएस ट्री पद्धति का प्रयोग करते हुए एक न्यू डेल्फी पद्धति विकसित की गई। परिदृश्य सृजित करने हेतु एक नया काल श्रृंखला अंतरक्षेप मॉडल उपनति प्रभाव विश्लेषण (टीआईए) किया गया। परिदृश्य सृजित करने हेतु डेल्फी एवं उपनति प्रभाव विश्लेषण को संयोजित किया गया। प्रस्तावित टीआईए के कार्यान्वयन के लिए एसएस मैक्रो विकसित किया गया।

- द्वि-स्तरीय प्रतिचयन अभिकल्पना के अंतर्गत अंशांकन पद्धति का प्रयोग करते हुए इस अवधारणा के साथ कि सहायक चर व्युत्क्रम रूप से अध्ययन चर से संबंधित हैं या नहीं, समष्टि प्राचलों के आकलन के लिए पद्धति विकसित की गई। कुल समष्टि तथा उनके अनुमानित प्रसरण के लिए तीन भिन्न उत्पाद विशिष्ट अंशांकन आकलक विकसित किए गए। विकसित उत्पाद विशिष्ट अंशांकन आकलकों के सांख्यिकी निष्पादन का आनुभविक आकलन करने हेतु एक अनुकार अध्ययन किया गया। कुल समष्टि के प्रस्तावित उत्पाद विशिष्ट अंशांकन आकलन का निष्पादन तब श्रेष्ठ पाया गया जब समष्टि में सभी पीएसयू के लिए समष्टि स्तरीय पूर्ण सहायक सूचना एसएसयू स्तर पर उपलब्ध होती है।
- व्यापक सर्वेक्षणों के तहत डाटा संग्रहण में सुधार लाने हेतु एक एंड्रोइड अनुप्रयोग नामक मोबाइल आधारित निजी साक्षात्कार (एमएपीआई) विकसित किया गया और एंड्रोइड टेबलेट्स का प्रयोग करते हुए उत्तर प्रदेश राज्य के दो जिलों तथा गुजरात राज्य के एक जिले के डाटा संग्रहण के लिए एमएपीआई उपलब्ध कराया गया। डाटा की समयबद्धता, यथार्थता और विश्वसनीयता के संदर्भ में, एमएपीआई को अधिक वर्तमान कागज आधारित सर्वेक्षण विधि की तुलना में लाभकारी पाया गया। सर्वेक्षण के स्थान से संबंधित प्रश्नों के साथ एमएपीआई सॉफ्टवेयर विकसित किया गया, जो स्थान की तस्वीरों के साथ सर्वेक्षण के जीपीएस स्थल को रिकॉर्ड करता है, जिसके कारण वर्तमान सर्वेक्षण विधि की तुलना में डाटा और अधिक विश्वसनीय हो जाता है। एमएपीआई सॉफ्टवेयर का ऑफलाइन वर्जन प्रतिदर्श सर्वेक्षण संसाधन सर्वर (<http://sample.iasri.res.in/ssrs/android.html>) पर ऑनलाइन उपलब्ध कराया गया है। अन्य विशिष्ट सर्वेक्षणों के लिए विकसित सॉफ्टवेयर को पंजीकृत उपयोगकर्ताओं के अनुरोध के आधार पर संशोधित किया जा सकता है। एमएपीआई सॉफ्टवेयर का ऑनलाइन वर्जन भी



विकसित किया गया, जो उपयोग करने हेतु उपलब्ध है।

- एफएओ (संयुक्त राष्ट्र का खाद्य एवं कृषि संगठन) प्रायोजित परियोजना के अंतर्गत विभिन्न देशों में अनेक वर्तमान स्थितियों के लिए मिश्रित एवं निरंतर फसलीकरण के तहत फसल क्षेत्र एवं उपज के आकलन हेतु प्रतिचयन पद्धतियाँ विकसित की गईं। विकसित पद्धति पर एफएओ द्वारा तीन निर्धारित देशों, एशिया, अफ्रीका और लेटिन अमेरिका/कैरिबियन प्रत्येक के एक क्षेत्र में, अर्थात् क्रमशः इंडोनेशिया, रवान्डा और जमैका में फील्ड में परीक्षण किया गया। संगणक आधारित निजी साक्षात्कार (सीएपीआई) का प्रयोग करते हुए इंडोनेशिया और जमैका और कागज आधारित निजी साक्षात्कार (पीएपीआई) का प्रयोग करते हुए सभी तीन देशों के दोनों जिलों में प्राथमिक डाटा संग्रहण किया गया। विकसित प्रतिचयन पद्धति को संयुक्त राष्ट्र के खाद्य एवं कृषि संगठन द्वारा भविष्य में वैश्विक रूप से, विशेषकर विकासशील देशों में अंगीकृत किए जाने की संभावना है। संयुक्त राष्ट्र के एफएओ द्वारा इस पद्धति को स्वीकार किया गया है और वैश्विक कार्यालय की तकनीकी रिपोर्ट के रूप में इसे प्रकाशित किया है। यह रिपोर्ट संयुक्त राष्ट्र वैश्विक कार्यनीति के एफएओ की वेबसाइट पर उपलब्ध है।
- बागवानी फसलों के क्षेत्र और उत्पादन के आकलन के लिए एक नई प्रतिचयन पद्धति, जो दक्ष, सरल, किफायती तथा समय की बर्बादी को रोकने में सक्षम है, का देश के छः राज्यों, यानी की महाराष्ट्र, तमिलनाडु, आंध्र प्रदेश, हिमाचल प्रदेश, हरियाणा और मध्य प्रदेश में वैधीकरण किया जा रहा है। विकसित पद्धति के वैधीकरण के पश्चात इसे बागवानी विभाग, कृषि सहकारिता और किसान कल्याण विभाग (डीएसीएफडब्ल्यू), कृषि और किसान कल्याण मंत्रालय, भारत सरकार द्वारा 2018-19 से देश के सभी राज्यों में कार्यान्वित किया जाएगा।

- खाद्यान्न काल-श्रृंखला डाटा की मॉडलिंग के लिए इष्टतम अपेक्षित बैंडविड्थ के आकलन की विभिन्न कार्यविधियों के निष्पादन का अध्ययन किया गया। वास्तविक खाद्यान्न उत्पादन एवं समेकित विकास दर मॉडलिंग की केरनल समुथिंग पद्धति के संबंध में, ईपानेकनिकोव केरनल (Epanechnikov kernel) श्रेष्ठ है। समेकित विकास दर के काल-श्रृंखला डाटा को सुदृढ़ करने हेतु लॉग मैमोरी त्रुटि प्रक्रम सृजित करने के लिए संगणक प्रोग्राम लिखा गया।
- भारत के विभिन्न क्षेत्रों में अनेक दलहनों के बाजार समेकन का अध्ययन किया गया। अध्ययन के परिणामों में यह दर्शाया गया है कि प्रमुख दलहनों के थोक तथा खुदरा मूल्यों में एक सबल सह-समेकन है, हालांकि सहसमेकन में विचलन है। कुछ दलहनों में सबल सह-समेकन है, जबकि कुछ में सह-समेकन अच्छा नहीं है। क्षैतिज सह-समेकन के अलावा, विभिन्न दलहनों के थोक एवं खुदरा मूल्यों के बीच उर्ध्व सह-समेकन अध्ययन भी किया गया। यह निष्कर्ष निकाला गया कि मूल्य संकेत समस्त क्षेत्रों तक संचारित होते हैं, जो इस बात का सूचक है कि एक क्षेत्र में मूल्य परिवर्तन अन्य क्षेत्रों में मूल्य परिवर्तनों से निरंतर रूप से संबंधित हैं, और अन्य क्षेत्रों के मूल्यों पर इनका प्रभाव पड़ता है। तथापि, मूल्य परिवर्तनों की दिशा और तीव्रता, दलहनों की मांग एवं आपूर्ति के बीच गत्यात्मक लिंकेज से प्रभावित हो सकती है।
- आंध्र प्रदेश और तैलंगाना राज्यों में फसल बीमा का प्रभाव मूल्यांकन किया गया। विश्लेषण में यह दर्शाया गया है कि अखिल भारतीय स्तर पर कुल 4 प्रतिशत फार्म परिवारों ने अपनी फसल का बीमा करवाया है, लेकिन चयनित राज्यों में फसल का बीमा कराए गए खेतिहर परिवारों का प्रतिशत 16.74 प्रतिशत है। फसल बीमा का लाभ लेने वाले अधिकतर बीमाकृत खेतिहर परिवार (98 प्रतिशत) उधार लेने वालों में से थे। अधिकतर (55.2 प्रतिशत) किसानों ने फसल बीमा का लाभ नहीं लेने के लिए, जागरूकता के अभाव को प्रमुख कारण बताया। कुल बीमाकृत किसानों में से केवल



9.23 प्रतिशत महिला किसानों ने अपनी फसल का बीमा करवाया था। कुल मिलाकर, गैर-बीमाकृत प्रतिचयनित किसान परिवारों में, बीमाकृत किसानों की तुलना में, कम साक्षरता थी। फार्म आउटपुट पर फसल बीमा के प्रभाव का मूल्यांकन करने हेतु प्रोपेनसिटी स्कोर मैचिंग (पीएसएम) विधि और इंस्ट्रूमेंटल वेरियेबल (आईवी) विधि का प्रयोग किया गया। दोनों विधियों ने यह दर्शाया कि फसल बीमा का फार्म इनपुट (कृषि उत्पादन) पर सकारात्मक एवं व्यापक प्रभाव है।

- प्रसंभाव्यता भिन्नात्मक समीकरण (एसडीई) विकसित किया गया और सप्रतिबंध पूर्वानुमान त्रुटि प्रसरण के साथ असंचारित डाटा का इष्टतम पूर्वानुमान प्राप्त करने हेतु उसका उपयोग किया गया।
- प्रणाली जीवविज्ञान अध्ययनों में जीन-जीन संबंध की मॉडलिंग के लिए, जीन विनियामक तंत्रिकाएँ दृश्यात्मक रूप से हमें विभिन्न जीवविज्ञानी प्रक्रमों का अध्ययन करने में सहायता प्रदान करती हैं। जीन व्यंजकता (जीई) डाटा में नॉइस एवं उच्च विमीयता ने अंतरनिहित जीन निनियामक अन्योन्यक्रियाओं का प्रतिरूपण करने हेतु कुछ सांख्यिकीय पद्धतियाँ विकसित की गईं। एक स्टेट स्पेस फ्रेमवर्क के तहत पार्टिकल फिल्टरिंग पद्धति का प्रयोग करते हुए नॉइसी माइक्रोऐरे डाटा से जीन्स के टू जीई मान प्राप्त करने हेतु आकलक प्राप्त किया गया। जीई मानों का प्रयोग करते हुए स्पारसिटी स्थिति के तहत जीआरएन का प्रतिरूपण एवं संरचना बनाने हेतु एक पद्धति विकसित की गई। जीन नेटवर्क में हब जीन्स (उच्च अन्योन्यक्रियाशील जीन) की पहचान करने के लिए एक पुनः प्रतिचयन आधारित पद्धति, हब जीन खोज विधि विकसित की गई। कंस वर्सिस कंट्रोल स्टडी के लिए निर्मित किए गए जीआरएन में हब जीन्स और उत्कृष्ट हब जीन्स की पहचान करने के लिए एक R पैकेज, डीएचजीए (भिन्नात्मक हब जीन विश्लेषण) भी विकसित किया गया। बिग क्रॉप जीई डाटा से

सूचनाप्रद जीन्स की पहचान करने हेतु वर्तमान एसवीएम-आरएफई कार्यविधि को सुदृढ़ कर एक उन्नत जीन चयन तकनीक का प्रस्ताव किया गया।

- पोजिशन-वार संरेखित अनुक्रम डाटासेट्स में पोजिशनों के परस्पर संबंध की खोज करने हेतु पद्धति विकसित की गई।
- किसानों/उत्पादकों की दृष्टि से ऐसे जीनप्ररूपों का चयन वांछनीय है, जो पर्यावरणों की तुलना में स्थायी हैं और उच्च उपजशील हैं। उच्च उपज के साथ स्थायी जीनप्ररूप के चयन के लिए बहु मानदंड निर्णयन (एमसीडीएम) तकनीक का प्रयोग करते हुए बेहतर निष्पादन तथा मिश्रित स्थायित्व मापदंड के साथ स्थायी जीनप्ररूपों का चयन करने के लिए दो समकालिक चयन सूचकांक विकसित किए गए। उच्च विमीय जीई डाटा से जैविक रूप से संबद्ध जीन्स का चयन करने हेतु सांख्यिकीय पद्धतियाँ विकसित की गईं। संबद्धता का अधिकतमीकरण कर तथा अधिकता (रिडनडेंसी) का न्यूनतमीकरण कर जैविक रूप से संबद्ध जीन्स का चयन करने के लिए एक उन्नत जीन चयन तकनीक, बूट-एमआरएमआर विकसित की गई। जीन चयन तकनीक का प्रयोग करते हुए चयनित प्राप्त जीन्स का वैधीकरण करने हेतु क्यूटीएल के साथ एक जीन सेट एनरिचमेंट टेस्ट (जीएसईटी) का प्रस्ताव किया गया। माइक्रोऐरे डाटा से नॉइस हटाने के लिए वेवलेंट विश्लेषण के आधार पर टू जीई मान प्राप्त करने हेतु एक आकलक विकसित किया गया।
- चावल में सूखा एवं लवण अपचयक जीन्स के लिए डाटाबेस विकसित किया गया। भारतीय बकरियों में आर्थिक रूप से महत्वपूर्ण गुणों से संबद्ध एसएनपी पर एक सूचना प्रबंधन प्रणाली विकसित की गई।
- संस्थान के सर्वर पर एनजीएस विश्लेषण तथा व्याख्या के लिए "भिन्नात्मक व्यंजकता व्याख्या एवं ट्रांसक्रिप्टोम असेम्बली के लिए उच्च निष्पादनीय संगणना" (hpcDATA) नामक एक वेब



आधारित सॉफ्टवेयर उपलब्ध कराया गया, जो टेस्टिंग के लिए एनएन पर उपलब्ध है (URL: <http://192.168.7.45:8084/hpcDATA>)। विकसित सॉफ्टवेयर ट्रांसक्रिप्टोम अनुक्रमण डाटा की डि-नोवो असेम्बली के लिए सक्षम है। इसके वर्कफ्लो में एसआरए टूलकिट, FastQC, ट्राइमोमेटिक, ट्राइनिटी एवं Rsem टूल्स हैं। इन समस्त टूल्स में लगभग 100 प्राचलों को विशिष्ट रूप से टेस्ट किया गया और इस पाइपलाइन में शामिल किया गया। इसे ट्रांसक्रिप्टोम डाटा के संदर्भ आधारित विश्लेषण (वर्कफ्लो) के लिए भी उपयोग किया जा सकता है, जिसमें एसआरए टूलकिट, FastQC, ट्राइमोमेटिक बोवाइट, टोफाट, कफलिंग एवं कफडिफ जैसे टूल्स पाइपलाइन में शामिल हैं। इसमें "भिन्नात्मक जीन अभिनिर्धारण जाँच मॉडल", "व्याख्या मॉड्यूल" एवं "परियोजना निगरानी मॉड्यूल" भी शामिल हैं।

- ऑपन वेब संसाधन (<http://webtom.cabgrid.res.in/ogr/>) विकसित किया गया, जिसे निःशुल्क ऐक्सस किया जा सकता है और इसमें सार्वजनिक डोमेन से उपलब्ध प्याज ईएसटी की असेम्बली की पूर्ण सूचना और ऐलियुम सेपा के ट्रांसक्रिप्टोम डाटा तथा उनकी व्याख्याओं एवं कार्यात्मक महत्ता के बारे में सूचना उपलब्ध कराई गई है।
- DIRPot: गैर-प्रतिरोधी प्रोटीनों से कीटनाशक प्रतिरोधी प्रोटीनों का विभेद करने के लिए एक वेब सर्वर <http://cabgrid.res.in:8080/dirprot/> विकसित किया गया और इसे गैर-प्रतिरोधी प्रोटीनों से कीटनाशक प्रोटीनों का विभेद करने के लिए उपयोग किया जा सकता है। यहाँ चार प्रकार के प्रतिरोधी प्रोटीनों पर विचार किया गया, जो कीटनाशक प्रतिरोधी जीन्स की चार श्रेणियों, अर्थात् साइटाक्रोम P450, एसिटाइल कोलिनेसटेरेस (AChE), नॉक डाउन प्रतिरोध (केडीआर), डाइलड्रिन से प्रतिरोध (आरडी) द्वारा कोडित है। सर्वर को डीपीसी विशिष्टताओं, अर्थात् डाइ-पेप्टाइड कम्पोजिशन (डीपीसी) को प्रशिक्षित किया गया है। इसके अलावा, पूर्वानुमान

के प्रयोजन हेतु बैकग्राउंड में आरबीएफ केरनल के साथ एसवीएम मशीन लर्निंग तकनीकों का उपयोग किया जाता है।

- R पैकेज WaveLetLongMemory विकसित किया गया और CRAN में सफलतापूर्वक अपलोड किया गया।
- डीएनए बारकोड का प्रयोग करते हुए प्रजातियों की पहचान करने के लिए एक संगणनात्मक विधि विकसित की गई। इस अवधारणा के आधार पर, वर्गिकी वैज्ञानिकों द्वारा प्रजातियों की सरलता से पहचान करने के लिए एक वेब सर्वर SPIDBAR (<http://cabgrid.res.in:8080/spidbar/>) भी विकसित किया गया।
- मीथिलेशन मीडियेटेड सबस्ट्रटयूशन और तात्कालिक व्याख्या के संबंध में प्राप्त अनुक्रमण विशिष्टताओं के आधार पर, इन्ट्रॉनिक अनुक्रमण से सीडीएस अनुक्रमण के विभेदन के लिए एक संगणनात्मक टूल DCDNC (<http://cabgrid.res.in:8080/DCDNC/>) विकसित किया गया। इस टूल को गोपशु एवं चावल डाटासेट में वैधीकृत किया गया। इस सर्वर का प्रयोग कर उपयोगकर्ता प्रायिकता के आधार पर, जिसके साथ प्रस्तुत टेस्ट अनुक्रमण का पूर्वानुमान कोडिंग अनुक्रमण के रूप में किया जाता है, आउटपुट प्राप्त करेगा।
- निश्चित प्रायिकता के साथ गैर-जीवाण्विक, गैर-विषाणु या गैर-कवक के रूप में कंडिडेट पेप्टाइड की प्रोपेनसिटी का पूर्वानुमान करने के लिए एक वेब सर्वर iAMPpred (<http://cabgrid.res.in:8080/amppred/>) विकसित किया गया। इस सर्वर के निष्पादन को वर्तमान टूल्स (AntiBP, CAMP2, AntiBP23, AVPPred4, iAMP-2L, ClassAMP6 and EFCFCBF7), जिन्हें गैर-जीवाण्विक पेप्टाइडों के पूर्वानुमान के लिए उपयोग किया जाता है, से तुलनीय पाया गया। इस सर्वर से पूर्वानुमानित गैर-जीवाण्विक पेप्टाइड, पेप्टाइड आधारित एंटीबाइटिकों की डिजाइनिंग में सहायता करेगा।
- उपयोगकर्ताओं के लिए एक R पैकेज, dhga



विकसित किया गया जिसे <https://CRAN.R-project.org/package=dhga> पर उपलब्ध कराया गया।

- उपयोगकर्ताओं के लिए एक R पैकेज, BootMRMR विकसित किया गया जिसे <https://CRAN.R-project.org/package=BootMR> पर उपलब्ध कराया गया।
- उपयोगकर्ताओं के लिए एक R पैकेज, GSAQ विकसित किया गया जिसे <https://CRAN.R-project.org/package=GSAQ> पर उपलब्ध कराया गया।
- GinmicrosatDB: तिल (सेसामुम इंडिकुम एल.) के लिए एक जीनोम-वार माइक्रोसेटलाइट मार्कर डाटाबेस विकसित किया गया (<http://backwin.cabgrid.res.in:8080/Gingelly7>). माइक्रोसेटलाइट मार्कर्स के विकास के लिए तिल किस्म 'स्वेता' के पूर्ण जीनोम अनुक्रमण का उपयोग किया गया ताकि मार्कर आधारित प्रजनन कार्यक्रमों में उपयोग हेतु पर्याप्त पॉलीमोर्फिक मार्कर्स की संख्या की उपलब्धता सुनिश्चित की जा सके।
- कृषि विज्ञान केंद्रों से किसानों तक ज्ञान एवं सूचना का प्रसार करने हेतु 'कृषि विज्ञान केंद्र ज्ञान नेटवर्क' नामक एक ऑनलाइन पोर्टल विकसित किया गया। इसे भाकृअनुप-भाकृसांअसं डाटा सेंटर पर होस्ट किया गया और यह URL (<http://kvk.icar.gov.in/>) पर उपलब्ध है। यह एक सिंगल विंडो प्लेटफॉर्म है, जो कृषि विज्ञान केंद्रों (केविके), कृषि एवं संबद्ध क्षेत्र से संबंधित कृषि विधियों के पैकेज के बारे में सूचना उपलब्ध कराता है, केविके को भी सुविधा प्रदान करता है, भारत के अधिकतर जिलों के लिए केविके द्वारा आयोजित घटनाक्रमों तथा कृषि आकस्मिकता योजना का विवरण उपलब्ध कराता है। सरल एवं उच्च सर्च सुविधाओं सहित पोर्टल में प्रश्न पूछने हेतु एक एकीकृत मॉड्यूल है। अटारी (ATTARI) एवं भाकृअनुप विस्तार प्रभाग के लिए पोर्टल में निगरानी सुविधा विकसित कर शामिल की गई

है। कृषि विज्ञान केंद्रों की प्रोफाइल, घटनाक्रमों, सुविधाओं और कृषि विधियों के पैकेज के लिए राज्य-वार एवं अटारी-वार निगरानी रिपोर्ट तैयार की गई। पोर्टल के माध्यम से केविके मासिक प्रगति रिपोर्ट (एमपीआर) और कृषि विस्तार एमपीआर (ईई-एमपीआर) रिपोर्टों को अपलोड कर सकते हैं, उन्हें देख सकते हैं तथा उनको अपडेट कर सकते हैं। उपयोगकर्ताओं को ट्रेक करने हेतु पोर्टल में एक लाइव क्लस्टर मानचित्र शामिल किया गया है। इस पोर्टल पर अब तक 6 लाख से अधिक उपयोगकर्ताओं ने दौरा किया है। पोर्टल के विभिन्न घटकों के लिए एक डैशबोर्ड निगरानी प्रणाली विकसित की गई।

- किसानों के लिए एंड्रोइड प्लेटफॉर्म में एक केविके मोबाइल ऐप विकसित किया गया। ऐप गुगल प्ले स्टोर पर उपलब्ध है और इसमें निम्नलिखित के बारे में सूचना उपलब्ध कराई गई है : केविके, सुविधाएं, कृषि विधियों, प्रश्न भेजें, आगामी घटनाक्रम, पिछले घटनाक्रम, मौसम एडवाइजरी, बाजार, चेंज केविके और केविके पोर्टल। किसान केविके में विशेषज्ञों से फार्म संबंधी कोई भी प्रश्न पूछ सकते हैं और इस ऐप के माध्यम से अपने प्रश्न का समाधान प्राप्त कर सकते हैं। सभी केविके प्रमुखों के लिए भाकृअनुप ई-मेल आईडी सृजित की गई। पोर्टल के उपयोगकर्ताओं को सहायता देने के लिए एक अलग मेलबॉक्स सृजित किया गया।
- भाकृअनुप मुख्यालय और कृषि विज्ञान चयन बोर्ड सहित भाकृअनुप के 111 संस्थानों में भाकृअनुप-ईआरपी का कार्यान्वयन किया गया। भाकृअनुप-ईआरपी के प्रभावकारी कार्यान्वयन के लिए पाँच स्थलों पर एचआरएमएस एवं पे-रोल तथा एससीएम मॉड्यूलों पर छः ज्ञान संवर्धन प्रशिक्षण कार्यक्रम आयोजित किए गए। भाकृसांअसं में भाकृअनुप-ईआरपी पर संस्थान विशिष्ट प्रशिक्षण कार्यक्रम आयोजित किए गए। सातवे वेतन आयोग की सिफारिश के अनुसार, पे-रोल मॉड्यूल को सफलतापूर्वक कस्टमाइज किया गया और ईआरपी



प्रणाली के माध्यम से 47 संस्थानों द्वारा वेतन संवितरित किया गया। रोकड़ आधार पर व्यय रिपोर्ट को कस्टमाइज एवं टेस्ट कर भाकृअनुप मुख्यालय के वित्त प्रभाग को प्रदर्शित किया गया।

- MS-PGE प्रणाली के मॉड्यूल के रूप में भाकृअसं की पीएच. डी. प्रवेश परीक्षा के लिए ऑनलाइन अनुप्रयोग प्रणाली विकसित की गई। वर्ष 2017 में इसकी शुरुआत से लगभग 3300 छात्रों ने स्वयं को इस प्रणाली के अंतर्गत अपना पंजीकरण कराया है। परीक्षा केंद्र के आबंटन केपीजी स्कूल की आवश्यकताओं के अनुसार, विभिन्न रिपोर्टें सृजित की गईं। इस प्रणाली में अभ्यर्थियों को आवेदन भरने के संदर्भ में, विभिन्न उपयोगकर्ताओं के प्रश्नों का समाधान करने तथा छात्र पंजीकरण एवं परिणाम को अपलोड करने संबंधी अन्य मॉड्यूलों के लिए सहायता उपलब्ध कराई गई है।
- सीआईएफई, मुंबई के लिए शैक्षणिक प्रबंधन प्रणाली विकसित की गई। प्रणाली के प्रोडक्शन इन्स्टांस (एनडीआरआई शैक्षणिक प्रबंधन प्रणाली) को स्थापित किया गया और मास्टर डाटाबेस विकसित करने का कार्य आरंभ किया गया। प्रणाली के प्रोडक्शन इन्स्टांस (आईवीआरआई शैक्षणिक प्रबंधन प्रणाली) को स्थापित किया गया और मास्टर डाटाबेस विकसित करने का कार्य आरंभ किया गया।
- वैज्ञानिक स्टाफ की संवर्ग पदस्थिति और स्थानांतरण के प्रबंध हेतु कार्मिक प्रबंधन प्रणाली

विकसित की गई जिसे भाकृअनुप में कार्यान्वित किया गया। भाकृअनुप में रिजल्ट फ्रेमवर्क डॉक्यूमेंट प्रबंधन प्रणाली पर वेब आधारित अनुप्रयोग की अभिकल्पना एवं विकास कार्य को पूरा कर लिया गया है।

- भाकृअनुप के शिक्षा प्रभाग के लिए भाकृसांस द्वारा विकसित वर्तमान समस्त समाधानों को भाकृअनुप के कृषि शिक्षा पोर्टल के तत्वावधान के अंतर्गत समाकलित किया गया है। छात्र पोर्टल विकसित करने जैसी अतिरिक्त नई पहलें आरंभ की गईं। एक वेब प्रतिक्रियात्मक होमपेज विकसित किया गया जिसे <http://education.icar.gov.in> पर उपलब्ध कराया गया। पोर्टल में शामिल की गई अन्य अतिरिक्त सुविधाओं में निम्नलिखित शामिल हैं: प्रत्येक घटक के एसओई को अपलोड करने के लिए बजट मॉड्यूल के तहत व्यय का विवरण (एसओई) उप.मॉड्यूल; प्रत्येक घटक के तहत मांग को अपलोड करने के लिए निधियों की मांग एवं दिशानिर्देश उप.मॉड्यूल, जो विश्वविद्यालय नोडल अधिकारी द्वारा विभिन्न शीर्षकों के तहत निधियों एवं प्रत्येक मॉड्यूल के लिए निधियों की मांग को ट्रैक करने हेतु अनेक नई रिपोर्टों की फाइलिंग में तथा शिक्षा प्रभाग द्वारा निधियाँ जारी करने में सुविधा प्रदान करता है और इन्हें इंटरनेट एक्सप्लोरर, गुगल क्रॉम और मोजिला फायरफॉक्स में देखा जा सकता है। मोबाइल फोन्स/टेबलेट्स पर देखने के लिए अनेक रिपोर्टों को वेब अनुक्रियाशील बनाया गया है।

भारत में कश्मीर के पर्वतीय कुपवाड़ा जनपद में प्रचलित कृषिवानिकी पद्धतियों के द्वारा संवित कार्बन स्टॉक तथा कार्बन अवशोषण क्षमता का आंकलन

अजीत, अनिल कुमार, ब्रह्मजीत गहलौत, विजय पाल सिंह, नरेश कुमार एवं उमा

सार

भारत की कश्मीर घाटी में स्थित कुपवाड़ा जनपद के वन क्षेत्र के बाहर ग्रीन हाऊस गैसों के शमन एवं कार्बन स्टॉक के आंकलन हेतु एक गत्यात्मक कार्बन अकाउन्टिंग मॉडल, CO₂FIX, का प्रयोग किया गया। प्राथमिक सर्वेक्षण से यह पता चला कि किसान के एक हेक्टेयर क्षेत्रफल के खेत में औसतन 135 पेड़ थे जिनमें मेलस (33.75%) पापुलस (29.91%) सैलिक्स (14.32%) जगलैन्स (6.68%) तथा रोबिनिया (4.7%) आदि पेड़ों की प्रजातियां प्रमुख थी। धान एवं मक्का खरीफ की प्रमुख फसलें हैं जबकि रबी के मौसम में तिलहन एवं चारा फसलों की प्रधानता रहती है। कृषिवानिकी पद्धतियों के सभी तीन कार्बन घटकों को एक साथ (जैसे पेड़, फसल एवं मिट्टी) लेकर कार्बन अवशोषण क्षमता के पूर्वानुमान का आंकलन 0.88 टन कार्बन/हेक्टेयर/वर्ष किया गया है। जनपद स्तर पर कृषिवानिकी पद्धतियों द्वारा 146,996 टन कार्बन डाई आक्साईड के समतुल्य प्रति वर्ष अवशोषण का अनुमान है, जो जम्मू एवं कश्मीर राज्य के प्रत्येक हिस्से में कृषि/सिंचाई कार्यों में विद्युत उपयोग के कारण उत्पन्न ग्रीन हाऊस गैसों के उत्सर्जन को पूर्णतया शमन कर सकता है।

शब्द व्याख्या—कृषिवानिकी पद्धतियों, कार्बन अवशोषण क्षमता, ग्रीन हाऊस गैस शमन, मृदा कार्बन, वृक्ष जैव-भार

विविधीकरण अवसर, उन्नत मृदा उपयोग दक्षता, सूक्ष्म जलवायु परिवर्तन, स्थायी सुरक्षा का प्रबन्ध आदि कृषि वानिकी व्यवस्था के कुछ ऐसे महत्वपूर्ण तत्व हैं जो जलवायु परिवर्तन के साथ अनुकूलन में आवश्यक भूमिका अदा करते हैं। यह प्रक्षिप्त किया गया है कि कृषि वानिकी के अन्तर्गत आने वाला वैश्विक क्षेत्रफल

लगभग 400 लाख हेक्टेयर है। क्योटो प्रोटोकॉल में कृषिवानिकी को, खनिज ईंधन प्रतिस्थापन के माध्यम से, वातावरणीय कार्बन डाई ऑक्साईड सघनता को कम करने के लिए एक विश्वसनीय पद्धति के रूप में मान्यता दी गई है। कश्मीर घाटी में साधारणतया वृक्ष आधारित कृषि प्रणाली को प्रमुखता से उपयोग में लाया जाता है। तथापि जम्मू कश्मीर प्रदेश में जिला स्तर पर कृषकों के खेतों के लिए वर्तमान व्यवस्था में कार्बन अवशोषण का कोई अनुमान उपलब्ध नहीं है। इस अध्ययन का उद्देश्य, जम्मू कश्मीर राज्य के कुपवाड़ा जनपद में किसानों के खेतों पर उपलब्ध में कृषिवानिकी पद्धति के माध्यम से कार्बन स्टॉक का मूल्यांकन करना तथा कार्बन अवशोषण सामर्थ्य का पता लगाना है। बहुत से अध्ययनों के द्वारा कश्मीर घाटी में पेड़ों की विभिन्न प्रजातियों के अन्तर्गत अवशोषित कार्बन स्टॉक पर रिपोर्ट दी गई है। हालांकि ये अध्ययन सीडस देवदार, फ्रैंकिजनस फ्लोरिबंडा एवं अल्मस बेलिचियाना जैसी कुछ चुनी हुई प्रजातियों तक ही सीमित रहे इसके अतिरिक्त ये अध्ययन फसलों, मृदा, चरागाह आदि के योगदान पर समग्र रूप से विचार नहीं करते हैं। कश्मीर घाटी के लिए किया गया यह शायद पहला अध्ययन है जिसके द्वारा किसानों के खेतों में प्रचलित कृषिवानिकी पद्धति के माध्यम से कार्बन अवशोषण का अध्ययन किया गया है। इसका उद्देश्य किसानों के खेतों पर कृषिवानिकी के सभी तीनों पूल जैसे वृक्ष, फसल तथा मृदा के कार्बन स्टॉक का एक साथ मूल्यांकन करना तथा प्रचलित पद्धति की कार्बन अवशोषण सामर्थ्य का आंकलन करना है। साथ ही इस अध्ययन का उद्देश्य जम्मू कश्मीर के कुपवाड़ा जिले के किसानों के खेतों पर अपनायी गयी कृषिवानिकी पद्धति की ग्रीन हाऊस



गैसों की शमन क्षमता का आकलन करना भी है।

विस्तृत ग्रामीण स्तरीय फील्ड सर्वेक्षणों, मृदा नमूने, जनपद स्तरीय जलवायु आकड़ों तथा एक प्रोसेस आधारित सिमुलेशन मॉडल आदि के संयोजन (मिश्रण) को प्रयोग करते हुए इस अध्ययन को सन 2014 में प्रारम्भ किया गया था। किसानों द्वारा उगाई जाने वाले पेड़ों की विभिन्न प्रजातियों की संख्या, पेड़ों की संख्या, डी बी एच (डायमीटर एट ब्रेस्ट हाईट या वक्ष स्तरीय व्यास), पेड़ की कुल ऊंचाई आदि का आकलन करने के लिए, ए आई सी आर पी ए एफ-श्रीनगर-क्रेन्द्र, जम्मू कश्मीर की टीम द्वारा एक सघन प्राथमिक सर्वेक्षण का संचालन मानक कार्य पद्धति अपना कर किया गया। पेड़ों की संख्या की गणना प्राकृतिक जंगल के बाहर के क्षेत्र में जनपद के सभी प्रखण्डों में की गई। बेसलाईन कार्बन तथा कृषिवानिकी पद्धति में संचित कार्बन का आकलन के लिए गत्यात्मक कार्बन अकाउन्टिंग मॉडल CO2FIX-3.1 का प्रयोग किया गया है। CO2FIX-3.1 मॉडल एक कार्बन आकलन मॉडल है जिसको CASFOR-II परियोजना के अन्तर्गत विकसित किया गया है। (dataservices.efi.int/casfor/project.html). इस मॉडल में कार्बन क्रेडिट तथा जैव भार को हेक्टेयर स्तर पर एक-वर्षीय दर के साथ सिमुलेट किया गया है। स्टैम-सी ए आई (तने की वार्षिक वृद्धि दर) की परिभाषा के अन्तर्गत परिमितीय विकास (आयतनी विकास) इस मॉडल का आधार भूत इनपुट है। शाखाएं, पत्तियां तथा जड़ें जैसे पेड़ के अन्य घटक, तने की वृद्धि के प्रतिशत के इनपुट के रूप में होते हैं। तने की आयतनीय वृद्धि भार (जैवभार) माड्यूल के माध्यम से जैविकभार वार्षिक वृद्धि में परिवर्तित हो जाती है। जैविकभार का मृदा में प्रवाह हार्वेस्ट पैरामीटरों तथा कुल विक्रय के भावों के माध्यम से संचालित होता है। इस मॉडल का प्रयोग कई अध्ययनों तथा विभिन्न परिस्थिति तन्त्रों में, क्योटो प्रोटोकॉल के अन्तर्गत कार्बन स्टॉक के मूल्यांकन के लिए किया गया है। भारत के गंगा क्षेत्रों में स्थित तथा प्रचलित कृषिवानिकी की कुछ चुनी हुई प्रजातियों की विलगन क्षमता तथा कार्बन भण्डारन के आकलन हेतु CO2FIX मॉडल का प्रयोग किया गया है।

इस मॉडल को <http://www.efi.int/project/casfor/co2fix/registewr32.php> से निशुल्क डाउनलोड किया जा सकता है तथा इस मॉडल की विस्तृत जानकारी <http://www.efi.int/projects/casfor> से प्राप्त की जा सकती है। CO2FIX का प्रयोग सेन्दुरी, रौध तथा प्रोकॉम्प जैसे मॉडलों, जो कि सामान्यतः एक प्रजाति को डील करते हैं, की तुलना में एक से अधिक प्रजातियों को एक साथ डील करने के लिए किया जा सकता है। पाठक (अजीत इत्यादि 2013, 2016) को विस्तृत विधियों इनपुट पैरामीटरों तथा CO2FIX कोहोर्ट्स जैसे पेड़, फसल व मृदा के पैरामीटराईजेशन के विषय में विस्तृत जानकारी हेतु देख सकते हैं। अवशोषण क्षमता तथा ग्रीन हाउस गैस शमन के आकलन के लिए आकलित वृक्षों की कुल संख्या को जनपद के कुल फसलीकृत क्षेत्रफल से विभाजित किया जाता है। इसके पश्चात् शमन क्षमता को दर्शाने के लिए वृक्षों के कार्बन अवशोषण क्षमता के आकलित मूल्य को कार्बन डाई आक्साईड के समतुल्य व्यक्त किया जाता है। कुपवाड़ा, कश्मीर घाटी का एक पिछड़ा सीमान्त जिला है तथा समुद्र तल से 5300 मीटर की ऊंचाई पर 34.17 तथा 34.12 डिग्री उत्तर के बीच तथा 73.16 डिग्री पूर्व में स्थित है। इसमें 11 खण्ड हैं जिनमें 363 गांव तथा 90418 के लगभग घर हैं। इस जिले की औसत वार्षिक वर्षा 1061 मिली मीटर है। नवम्बर के महीने में वर्षा का स्तर सबसे कम यानि 28 मि. मी. होता है। जुलाई, वर्ष का सबसे गर्म महीना होता है। जबकि औसत तापमान 23.9°C रहता है। जनवरी का औसत तापमान 2.9°C रहता है। जोकि वर्ष का निम्नतम औसत तापमान स्तर है। अधिकतम तथा न्यूनतम वर्षा (वृष्टि) का अन्तर लगभग 119 मि. मी. रहता है। एकत्रित आकड़ों से अनुमान लगाया गया कि विभिन्न कृषि वानिकी प्रणालियों, खेतों के अन्दर व किनारों पर, छोटे भूखण्डों, बागवानी उद्यानों, परती भूमि, नालों आदि पर उगे पाये गये वृक्षों की संख्या लगभग 6,198,349 थी।

कुपवाड़ा जिले में पाये गये वृक्षों की विभिन्न प्रजातियों की औसत संख्या 26 रिकार्ड की गई। बढ़वार की गति के आधार पर 9 प्रजातियों को तीव्र, 10 को मध्यम तथा 7 को मन्द वर्गों में रखा गया। कुल



मिलाकर 472,863 वृक्षों को मन्द गति से बढ़ने वाले वृक्षों की श्रेणी में रखा गया। इस श्रेणी में आने वाले वृक्षों में अखरोट सबसे अधिक संख्या (414274) वाले थे। इसके बाद क्रमशः सेल्टिस, देवदार, कैल, अंजीर, चिनार तथा बलूत के वृक्ष इस श्रेणी में पाये गये। सर्वेक्षण में पाया गया कि वृक्षों की अधिकतम औसत ऊँचाई 22.28 मीटर थी जबकि अधिकतम औसत डी बीएच (वृक्ष स्तरीय व्यास) 1.53 मीटर थी। मध्यम श्रेणी के वृक्षों की संख्या 2402420 थी जिनमें सबसे अधिक संख्या (2092262) सेब के वृक्षों की थी। उसके बाद आने वालों में क्रमशः शहतूत, नाशपाती, सनौवर, नीम्बू, सन्तरा तथा बादाम के वृक्ष रहे हैं। आकड़ों से यह भी पता चला कि अलभस प्रजाति ने 13.79 मीटर की अधिकतम औसत ऊँचाई प्राप्त की तथा मोरस प्रजाति ने 0.34 मीटर की अधिकतम व्यास प्राप्त की। मध्यम गति की बढ़वार वाली प्रजातियों में नीम्बू व संतरे छः खण्डों में पाई गयी तथा इस श्रेणी की शेष सभी प्रजातियाँ जिले के सभी खण्डों में सामान्यतः सभी जगहों पर पायी गईं।

जिले में तीव्र गति की बढ़वार वाली प्रजातियों की श्रेणी के 3326065 वृक्षों में पोपलर 1854288 वृक्षों के साथ प्रथम स्थान पर था उसके बाद क्रमशः आने वालों में मैलिक्स, रोबिनिया, एलेन्यस, पीछ (आड़ु), आनार, खुमानी (बंगली) तथा चेरी रहे। अधिकतम औसत ऊँचाई व व्यास की मात्रा 22.65 तथा 0.33 मीटर पोपलर वृक्ष के लिए रिकार्ड की गई। यह बात ध्यान आकर्षित करने वाली है कि जिले के सभी खण्डों में पोपलर तीव्र बढ़वार वाली श्रेणी में सर्वाधिक वरीयता वाला वृक्ष था।

किसान के खेतों में उपलब्ध प्रति हेक्टेयर औसत वृक्षों की संख्या 135.76 थी। बढ़वार के दृष्टिकोण से उच्च, मध्यम एवं निम्न श्रेणी में वृक्षों की औसत संख्या प्रति हेक्टेयर क्रमशः 72.79, 52.62 तथा 10.35 थी। उच्च श्रेणी वाली प्रजातियों में पोपलर 29.91% के साथ पहले स्थान पर रहा तथा इसके बाद 2.33% के साथ मोरस का नम्बर आता है। मन्द बढ़वार वाली श्रेणी की प्रजातियों में 6.68% के साथ जुगलान्स सबसे प्रबल प्रजाति रही। (तालिका -1)।

फसल उत्पादकता के दूसरे क्रम के आँकड़े एन आई सी व जिला प्रशासनिक कार्यालय, कुपवाड़ा से प्राप्त किये गये। अध्ययन में शामिल चार वर्ग समूहों में से तीन, वृक्षों में से शामिल किये गये जैसे तीव्र, मध्यम और मन्द जबकि चतुर्थ वर्ग समूह फसल का रखा गया। 6.2 तथा 2.0 टन प्रति हेक्टेयर की औसत उत्पादकता के साथ धान और मक्का की फसले कुपवाड़ा जिले की महत्वपूर्ण तथा लोकप्रिय खरीफ फसलों में रही। तिलहन एवं पशुचारे की फसले 0.6 तथा 17.5 टन प्रति हेक्टेयर की उत्पादकता के साथ रबी में उगायी जाने वाली फसलें रही। सेब व अखरोट दो सबसे अधिक वरीयता प्राप्त फसलों में रहीं जिनका औसत उत्पादन क्रमशः 8.5 व 2.3 टन प्रति हेक्टेयर था।

CO₂FIX माडल को फसल उत्पादकता के प्राथमिक सर्वेक्षण तथा NIC से प्राप्त द्वितीय क्रम के आकड़ों के साथ पैरामीटराइज्ड किया गया। अधिकतम व न्यूनतम तापमान, अवक्षेपण, वाष्पन उत्सर्जन आदि जिला स्तरीय जलवायु सम्बन्धी आँकड़े गत्यात्मक सिमुलेशन मॉडल तथा मृदा अपघटन YASOO सवरूटिन में फीड किये गये। इस मॉडल से यह पता चला की जिला स्तर पर किसानों के खेतों में स्थित वृक्षों की बेस लाईन में संग्रहित वृक्ष भार (बायोमास) 99.59 टन प्रति हेक्टेयर था तथा यह आकलित किया गया कि इसकी मात्रा सिम्यूलेशन की 30 वर्ष की अवधि में 147 टन प्रति हेक्टेयर तक पहुँच सकता है। जिला स्तर पर वृक्ष भार में वृद्धि की मात्रा 1.58 टन प्रति हेक्टेयर आकलित की गई है। प्राप्त परिणाम डार इत्यादि 2015 के अनुसंधान से मेल खाते हैं। जिन्होंने कश्मीर घाटी के समशीतोष्ण जंगलों, अन्डरस्टोरी, मृत काष्ठ, फलोर लिटर तथा मिट्टी के बायो मास व कार्बन स्टॉक का मापन किया तथा यह बताया कि वृक्ष बायोमास 100.8 से 294.8 टन प्रति हेक्टेयर के बीच रहा।

वेसलाइन में कुल बायोमास (वृक्ष व फसल) 102 टन प्रति हेक्टेयर था तथा इसके 149.85 टन प्रति हेक्टेयर पहुँचने की सम्भावना है। CO₂FIX मॉडल के परिणामों के सिमुलेशन यह पूर्वानुमान देते हैं। कि बायोमास कार्बन में 48.91 टन प्रति हेक्टेयर की बेस लाईन से 71.78 टन हेक्टेयर की वृद्धि हो जायेगी।



बैसलाईन मृदा कार्बन 22.28 टन प्रति हेक्टेयर था तथा ऐसा आंकलन है कि 30 वर्षों के सिमुलेशन की अवधि में यह 25.99 टन प्रति हेक्टेयर तक बढ़ेगा। जनपद स्तर पर मृदा कार्बन विलगन की दर 0.1236 टन प्रति हेक्टेयर आकलित की गई। अजीत इत्यादि 2013 ने सूचित किया कि गंगा के मैदानी क्षेत्र में स्थित 13 वृक्ष प्रति हेक्टेयर वाले सुल्तानपुर जनपद के लिए मृदा कार्बन विलगन की आकलित दर 0.016 टन प्रति हेक्टेयर थी तथा 24 वृक्ष प्रति हेक्टेयर वाले दीनाजपुर जनपद के लिए यही दर 0.37 टन प्रति हेक्टेयर थी। वास्तव में मृदा कार्बन विलगन कई घटकों पर निर्भर करता है जैसे प्रति हेक्टेयर क्षेत्रफल में उपस्थित वृक्ष, वृष्टि (वर्षा), तापमान, सूर्य की धूप की अवधि, तथा प्रबन्धन की गुणवत्ता के साथ-साथ स्थानीय जलवायु परिस्थितियाँ।

कुपवाड़ा जिले में अपनायी जा रही कृषिवानिकी प्रणाली की कुल कार्बन अवशोषण क्षमता 0.88 टन/हेक्ट./वर्ष आकलित की गई है जो राष्ट्रीय स्तर 0.21 टन/हेक्ट./वर्ष से तुलनात्मक रूप से अधिक है। वर्तमान में उजागर किये तथ्य जेना और यादव इत्यादि के परिणामों जैसे ही हैं जिन्होंने यह आकलन किया है कि कृषि वानिकी के अन्तर्गत आने वाला विश्व भर का वर्तमान क्षेत्रफल 400 मिलियन हेक्टेयर है जिसके अन्तर्गत 0.72 टन/हेक्ट./वर्ष कार्बन प्राप्त होता है।

जम्मू व कश्मीर में ग्रीन हाऊस गैसों का उत्सर्जन मुख्यतः खनिज ईंधन के जलाने के कारण होता है। विद्युत उत्पादन, जिसमें कि राज्य के पारम्परिक विद्युत उत्पादन करने वाले प्लांटों की अधिकतम हिस्सा है से भी महत्वपूर्ण स्तर पर ग्रीन हाऊस गैसों का उत्सर्जन होता है। भारत के पर्यावरण तथा बन मन्त्रालय द्वारा इस बात का मूल्यांकन किया गया कि जम्मू व कश्मीर राज्य में बिजली की खपत से ग्रीन हाऊस गैसों के उत्सर्जन में कितना योगदान है। (<http://www.wsef.nic.in/sites/default/files/space/janmu-kashmir.pdf>), जिसमें पाया गया कि कृषि/सिंचाई में बिजली की खपत से होने वाले ग्रीन हाऊस गैस उत्सर्जन 132229.80 टन कार्बन के समतुल्य था। सिमुलेशन के 30 वर्ष की अवधि पर विचार करने हुए वर्तमान अध्ययन के

आधार पर प्रचलित कृषिवानिकी प्रणाली की कार्बन अवशोषण क्षमता कुपवाड़ा जिले के लिए 0.88 टन/हेक्ट./वर्ष आकलित की गई है जिसके अनुसार वातावरण से कार्बन डाई आक्साईड विलगन का स्तर 3.23 टन/हेक्ट. आकलित किया गया। कुपवाड़ा जिले में कुल बोये गये 45651 हेक्टेयर क्षेत्रफल में जिला स्तर पर प्रचलित कृषिवानिकी प्रणाली से वार्षिक 146996.22 टन कार्बन अवशोषण का आकलन किया गया है। अर्थात् जिला स्तर पर वार्षिक 0.150 लाख टन कार्बन डाई आक्साईड के समतुल्य कुपवाड़ा जिले के किसानों के खेतों में प्रचलित कृषिवानिकी प्रणालियाँ से अवशोषण का आकलन है जो कि सम्पूर्ण जम्मू कश्मीर राज्य में कृषि व सिंचाई में होने वाले विद्युत खपत से होने वाले ग्रीन हाऊस गैसों के उत्सर्जन को पूर्णतया शमन कर देगी। इस प्रकार इस अध्ययन के आँकड़े, परिमाणत्मक रूप से किसानों के खेतों में अपनायी गयी कृषिवानिकी प्रणाली के माध्यम से पर्यावरण तन्त्र की ग्रीन हाऊस गैसों के शमन में योगदान तथा दूसरी अन्य भूमिका की पुष्टि करते हैं जैसे की उनकी खाद्यान्न, ईंधन, फल तथा पशुचारा आदि प्रदान करने की प्राथमिक भूमिका।

संदर्भ

1. अल्ब्रेट, ए. एवं कान्डजी, एस.टी., कार्बन सीक्वेश्ट्रेशन इन ट्रॉपिकल एग्रोफोरेस्ट्री सिस्टम, एग्रिक, इकोसिस, एनवायर्न., 2003,99,15-27.
2. शव, के.पी.सी., वेर्चोट, एल.वी.व., एवं लारमैन, जे., एडेप्टेशन टु क्लाइमेट चेंज थ्री सस्टेनेबल मैनेजमेंट एण्ड डवलपमेंट आफ एग्रोफोरेस्ट्री सिस्टम एस ए टी ईजर्नल, 2007 ,4,1-30.
3. शेख, ए. कर., रिकंदर, बी. एम., पंडित, ए.के. एवं गनई, बी.ए., टेरिस्ट्रीयल कार्बन सीक्वेश्ट्रेशन एज ए क्लाइमेट चेंज मिटिगेशन एक्टिविटी. जे. पोल्यूट. इफैक्ट्स कंट्रोल, 2014, 2 110; डी ओ आई : 104172/ जे पी ई.1000110.
4. कैलफैपिएट्रा, सी., जीलेन,बी., कर्नास्की, डी., स्यूलेमैस, आर. एवं मुग्नोज्जा, जी.एस, रैस्पॉस एण्ड पर्टेशियल आफ एग्रोफोरेस्ट्री क्राप्स अन्डर ग्लोबल चेंज. एनवायर्न. पौल्यूट., 2010,158,1095-1104.



5. वानी, एन. आर. इत्यादि, कार्बन परसेंट इन डिफरेंट कम्पोनेंट आफ ट्री स्वीसीज ऐण्ड सॉयल आर्गोनिक कार्बन पूल अन्डर दीज ट्री स्वीसीज इन कश्मीर वैली. कर वर्ल्ड एनवायर्न., 2014, 9 (1),17-181; <http://dx.doi.org/10.12944/cwe.9.1.24>.
6. वानी, एन.आर. इत्यादि, ग्रोथ परफेरेमेंस, बायोमास प्रोडक्शन ऐण्ड कार्बन स्टोक्स ऑफ 19 ईयर ओल्ड फ्रेक्सिनस फ्लोरिबन्डा (एश ट्री) प्लांटेशनस इन कश्मीर बैली, एग्रिक. फोरेस्ट., 2014,60(1), 125-143.
7. वानी, एन.आर. इत्यादि, बायोमास, कार्बन स्टॉक्स ऐण्ड कार्बन डाई आक्साईड मिटिगेशन पटोशिमल ऑफ सेड्स देवदार अण्डर टेम्परेट कन्डिशनस ऑफ कश्मीर. कैन. जे. प्योर एप्लि. सॉडि, 2014,8(1), 2677-2684.
8. नेबर्स, जी. एस. इत्यादि, कार्बन प्रॉफईल ऑफ टिपिकल फोरेस्ट टाईप्स एक्रीस यूरोप एसेस्ड विद CO2 FIX. इकोल. इंडिकेटर्स, 2002,1,213-233.
9. मसेरा, ओ. इत्यादि, माउलिंग कार्बन सीक्वेस्ट्रेशन इन एफकॉरेस्टेशन एग्रोकॉरेस्ट्री ऐण्ड फारेस्ट मैनेजमेंट प्रोजेक्ट्स: द CO2 FIX V.2 अप्रोच. इकोल. माउल, 2003,164, 177-199.
10. स्केलास, एन. जे. इत्यादि CO2 FIX V.3.1-1, माउलिंग फ्रेमवर्क फॉर क्वांटिफाईंग कार्बन सीक्वेस्ट्रेशन इन फारेस्ट ईकोसिस्टम्स. ALTERRA रिपोर्ट 1068, वेजनिंजेन, द नीदरलैंड, 2004.
11. गैबरी, एस. इत्यादि, एस्टिमेटिंग द नेट कार्बन बैलेस बोरियल ओपन बुडलैंड अफोरेस्टेशन: अ स्टडी इन क्यूबेक्स क्लोज्ड क्राऊन बोरियल फरिस्ट फरिस्ट इकोल. मैनेज, 2009,257. 483-494.
12. कौल, एम. इत्यादि, कार्बन स्टोरेज ऐण्ड सीक्वेस्ट्रेशन पटेंशियल ऑफ सलेक्टिड ट्री स्पीसीज इन इंडिया मिटिंग. एडेप्ट स्ट्रेटेज. ग्लोबल चेंज, 2010, 15,489-510.
13. अजीत इत्यादि माउलिंग एनेलिसिस ऑफ पटेंशियल कार्बन सीक्वेस्ट्रेशन अण्डर एक्जिस्टिंग एग्रोकॉरेस्ट्री सिस्टम इन श्री डिस्ट्रिक्ट्स ऑफ इंडो-गैंगेटिक प्लेन इन इंडिया एग्रोकॉरेस्ट सिस्ट, 2013,87,1129.1146.
14. अजीत इत्यादि एस्टिमेटिंग कार्बन सीक्वेस्ट्रेशन पोटेंशियल ऑफ एक्जिस्टिंग एग्रोकॉरेस्ट्री सिस्टम इन इंडिया एग्रोकॉरेस्ट सिस्ट, 2016,20: डीओ आई: 10.1007/&10457-016-9986-2 (पब्लिशड आनलाईन 12 अगस्त,2016).
15. लिस्क, जे. इत्यादि, कार्बन ऐण्ड डीकम्पोजिशन मॉडल yasoo फार फोरेस्ट स्वायल ईको, मॉडल, 2005,189,168-182.
16. डार, जे. ए. इत्यादि वेरियेशन ऑफ बायोमास ऐण्ड कार्बन पूल्स विद फोरेस्ट टाईप इन टेम्परेट फारेस्ट्स ऑफ कश्मीर हिमालया इंडिया एनवायर्न मोनितर एसेस, 2015,87(2), 55; डी ओ आईन 10.1067/510661-015-4299-7.
17. जाना, बी. के. इत्यादि, कार्बन सीक्वेस्ट्रेशन रेट ऐण्ड अबव ग्राऊन्ड बायोमास कार्बन पटोशियल ऑफ फोर यंग स्पीसीज जे. इकोल. नेचुरल. एनवायर्न 2009,1,15-24.
18. यादव ए. के. इत्यादि, बायोमास प्रोडक्शन ऐण्ड कार्बन सीक्वेस्ट्रेशन इन डिफरेंट एग्रोकॉरेस्ट्री सिस्टम इन तराई रीजन ऑफ सेंद्रल हिमालय इंडियन फॉरेस्टर, 2010,136, 234-244.



तालिका-1: प्राथमिक सर्वेक्षण में आकलित की गई वृक्ष प्रजातियों का प्रतिशत योगदान

मन्द गति से बढ़ने वाले वृक्ष	मध्यम गति से बढ़ने वाले वृक्ष	तीव्र गति से बढ़ने वाले वृक्ष
जुगलैन्स उपजाति (6.68)	मालस उपजाति (33.75)	प्राँप्लास उपजाति (99.59)
सिडरसो डोडरा (0.23)	मोरस उपजाति (2.33)	ऐहिन्यस एक्सलेसा (2.87)
फिक्स उपजाति (0.04)	परशुस उपजाति (1.68)	परुशु उपजाति (0.18)
क्वार्कस उपजाति (0.0005)	कपडास्ट्रोर लुसा (0.75)	सैलिक्स उपजाति (41.32)
सेतिरसास स्ट्रेला (0.37)	एस्क्यूअल इडिका (0.17)	रोबिनिफेडोसाचिया (4.7)
पाइनस पाविहियाना (0.21)	उत्मस उपजाति (0.03)	परुशु उपजाति (0.65)
प्लाटिनस उपजाति (0.02)	फ्रैक्सिनस उपजाति (0.011)	पूनिकग्रानैटन (0.56)
	साइट्रस उपजाति (0.004)	परुशु उपजाति (0.77)
	साइट्रस उपजाति (0.0038)	
	पुनास मैग्डस (0.003)	

तालिका-2: कश्मीर घाटी के कुपवाड़ा जिले में प्रचलित कृषिवानिकी पद्धतियों के अन्तर्गत वृक्ष/फसल घटकों में संचित जैवभार (CO₂FIX मॉडल का प्रयोग कर अनुकरण किये)

वृक्ष जैवभार (जमीन के ऊपर एवं नीचे) टन/हेक्ट	आधार रेखा जैवभार	99.59
	अनुकरण/सिमुलेटेड	147.19
कुल जैवभार (वृक्ष+फसल) टन/हेक्ट	आधार रेखा	102.18
	अनुकरण/सिमुलेटेड	149.85
मिट्टी-कार्बन (टन कार्बन/हेक्ट)	आधार रेखा कार्बन	22.28
	अनुकरण/सिमुलेटेड	25.99
जैवभार कार्बन (टन/हेक्ट)	आधार रेखा	48.91
	अनुकरण/सिमुलेटेड	71.78
कुल कार्बन (जैवभार+मिट्टी) (टन हेक्ट)	आधार रेखा	71.19
	अनुकरण/सिमुलेटेड	97.77
तीस वर्षों की अवधि में अनुकरण के समय एग्रोफोरेस्ट्री पद्धति में वास्तविक कार्बन को अलग किया (अन/हेक्ट)	कार्बन अलग	26.58
कृषिवानिकी पद्धति की अनुमानित वार्षिक कार्बन की अवशोषण क्षमता (टन/हेक्ट/वर्ष)		0.88
कृषिवानिकी पद्धति की अनुमानित Co ₂ शमन क्षमता (टन)		3.22
जिला स्तर पर कृषिवानिकी पद्धति की अनुमानित Co ₂ शमन क्षमता (Co ₂ की मिलियन टन वार्षिक समतुल्य)		0.15

बकरियों में वृद्धि वक्र प्राचलों की आनुवंशिक परिवर्तनशीलता : ब्रूट-स्ट्रेप तकनीकों का अनुप्रयोग

ए.के. पॉल, रंजित कुमार पॉल, लाल मोहन भर, हिमाद्रि शेखर राय, प्रकाश कुमार, सविता वधवा एवं सत्यपाल सिंह

सार

वृद्धि पशुओं के जीवन में एक महत्वपूर्ण चरण है, जो उत्पादन के विभिन्न रूपों को प्रभावित करती है जैसे दूध, मांस इत्यादि। भार-आयु आँकड़ों की विश्लेषणात्मक व्याख्या करना कठिन है अतः सांख्यिकीय रूप से जानवरों की वृद्धि का अध्ययन वांछनीय है। मूल्यांकन परिवर्तन के लिये और विकासशील प्रजनन योजनाओं के लिए वृद्धि वक्रों की आनुवंशिकता को समझना बहुत कठिन है। चयन प्रक्रिया के वृद्धि प्राचालकों की प्रायोगिक क्षमता का निरीक्षण करना आनुवंशिक प्राचालकों के लिए बहुत अनिवार्य है। वर्तमान अध्ययन में चार मॉडल्स लॉजिस्टिक, वॉन-बर्टलैनफी, गॉम्पर्टज और वीबुल को 142 पशुओं (बकरियाँ) के शारीरिक भार पर फिट किया गया।

विचाराधीन आँकड़ों के लिए सर्वोत्तम मॉडल वॉन-बर्टलैनफी है। अन्त में, वृद्धि वक्र प्राचालकों के सांख्यिकीय गुणों को ब्रूट-स्ट्रेप तकनीक का प्रयोग कर उन पर चर्चा की गई। आनुवंशिक प्राचालकों के वितरण को असामान्य पाया गया। परिपक्व भार और परिपक्व दर के बीच आनुवंशिक सहसंबंध ऋणात्मक सहसंबंध है जो दर्शाता है कि परिपक्वता दर के कारण हल्का परिपक्व वजन हो सकता है। परिपक्व भार की अधिक आनुवंशिकता यह दर्शाती है कि चयन प्रक्रिया के लिए परिपक्व भार का प्रयोग किया जा सकता है।

परिचय

विकास आनुवंशिक और पर्यावरणीय प्रभावों का आयोजन है। इसे प्रति इकाई समय में शरीर के आकार में बढ़ोतरी के रूप में परिभाषित किया जा सकता है। शरीर के वजन और आयु के संबंध विशेष रूप से मांस उत्पादन पशु, बकरी, सुअर, भेड़ आदि के

रूप में पशुओं में महत्वपूर्ण है। विकास अध्ययन में, आमतौर पर विकास मॉडल को फिट किया जाता है जो कि जैविक विकास की आवश्यक जानकारी प्रस्तुत करता है। चूंकि भार-आयु के आँकड़ों की व्याख्या विश्लेषणात्मक रूप से कठिन है, इसलिए सांख्यिकीय रूप से, पशुओं के विकास का अध्ययन किया जाता है तथा पशुओं के विकास आँकड़ों के सांख्यिकीय विश्लेषण में एक उपयुक्त वृद्धि वक्र मॉडल और प्रयोगों के समूह के परिणामों को एकत्रित करने का एक उचित विधि का चयन शामिल है। कुशल चयन प्रोग्रामों का विकास कर दोनों पौधे और पशुओं में आर्थिक महत्व के अधिकांश लक्षणों में सुधार प्राप्त किया जा सकता है। कुशल चयन करना तभी संभव है जब पॉपुलेशन के आनुवंशिक गुणों के साथ साथ लक्षणों के बारे में पूर्ण जानकारी प्राप्त हो। चयन और आनुवंशिक अग्रिम का परिणाम चयन की तीव्रता प्रारम्भिक पॉपुलेशन वयक्तिगत और चयन की सटीकता के बीच आनुवंशिकता के परिणाम पर निर्भर करता है। आनुवंशिक परिवर्तनशीलता का मात्रात्मक माप आनुवंशिक मापदंड (आनुवंशिकता) प्रजनन मूल्य को फीनोटाइपिक मान के रूप को मार्गदर्शन को रूप में विश्वनीयता को व्यक्त करती है। इसके आगे आनुवंशिकता न केवल लक्षणों के गुण को व्यक्त करती है किसी की पॉपुलेशन में एक लक्षण में सुधार अलगाव में नहीं किया जा सकता क्योंकि यह मुख्य लक्षणों में चयन के बाद पॉपुलेशन के अन्य सहसंबंधित लक्षणों को भी बदलता है इसलिए लक्षणों के बीच आनुवंशिक सहसंबंध की जानकारी रखने के लिये यह बहुत महत्वपूर्ण है।

उपयुक्त चयन सूचकांक के निर्माण के समय में दो या दो से अधिक लक्षणों में एक साथ सुधार करने



में भी बहुत उपयोगी है। ब्राउन इत्यादि (1976)ने मादा पशुओं में अरैखिक मॉडल को फिट किया। परिपक्व वजन और परिपक्वता दर के बीच लगातार नकारात्मक सहसंबंध था।

डी निस (1982) ने परिपक्व आकार और परिपक्व दर के बीच -0.57 के रूप में नकारात्मक आनुवंशिक सहसंबंध का अवलोकन किया।

कैचमैन इत्यादि (1988) ने अपने अध्ययन में पाया कि इनफ्लेक्शन के बिन्दु पर आयु दूसरे तीसरे सप्ताह की आयु पर शरीर के वजन के साथ मामूली रूप से सहसंबंध का परिपक्व शरीर के भार के साथ सहसंबंध नहीं था।

खेज इत्यादि (2002) में पीनस टाइडा के लिये वृद्धि वक्र आनुवंशिक प्राचलों का आकलन किया। उन्होंने अनंतस्पर्शी (asymptote) और प्राचल दर के बीच नकारात्मक सहसंबंध पाया जो दर्शाता है कि छोटे अनंतस्पर्शी के साथ वक्र शीघ्र ही अधिकतम ऊँचाई तक पहुँचते हैं। वाही, इत्यादि (2004) ने क्रॉसेज और शुद्ध भारतीय बकरियों के प्रजनन में वृद्धि प्रारूप पर अध्ययन किया। गोम्पर्टज माडल बकरियों के वृद्धि प्रारूप की व्याख्या करने में उत्तम मॉडल पाया गया। लॉजिस्टिक अरैखिक वृद्धि मॉडल वृद्धि वक्र के चरों के आनुवंशिक प्राचलों के आकलन करने के लिये फिट किया गया। लाम्बे इत्यादि (2006) ने जन्म से वध तक दो विपरीत नस्ल के भेड़ के बच्चों में वृद्धि की व्याख्या करने के लिये कई प्रयोग किये गये विभिन्न मॉडल की तुलना की और वृद्धि वक्र प्राचलों के आनुवंशिक प्राचल प्राप्त किये। कार इत्यादि (2006) का उद्देश्य विभिन्न अरैखिक वृद्धि मॉडल्स का प्रयोग कर 26 अकंसी (सफेद बकरी) मादा बकरी के बच्चे की वृद्धि का निर्धारण करना था।

हाल के वर्षों में पुनः प्रतिदर्श तकनीकों के आधार पर अधिक विश्वसनीय विधि को शामिल किया गया। इनमें से सबसे अधिक प्रचलित विधि जैकनाइफ (एफ्रान, 1979) और बूटस्ट्रेप विधि (एफ्रान, 1982) हैं। आस्टवीट (1990) ने पौधों के प्रजनन संबंधी आँकड़ों में प्रसरण और सह प्रसरण संबंधों के आनुवंशिक घटकों

के विश्वसनीय अंतराल प्राप्त करने के लिये बूटस्ट्रेप तकनीक को लागू किया। जयशंकर (1992, 1996) हेरिटिविलिटी के आकलन के लिये तथा हेरिटिविलिटी के गुणों का अध्ययन करने के लिये बूटस्ट्रेप तकनीक का प्रयोग किया। सिंह तथा वाही (2003) ने हाफ सिब माडल में हेरिटिविलिटी की मानक त्रुटि के आकलन के लिये इष्टतम परिवार आकार तथा संरचना प्राप्त की।

सिंह इत्यादि ने (2006) हाफ सिब मॉडल मास्टर प्रतिदर्श स्वतंत्र रेखाचित्र के द्वारा आनुवंशिकता के विचलन का आकलन बूटस्ट्रेप की रोबस्टनेस से किया। शरीर के वजन-आयु के आँकड़ों की एक श्रृंखला की विश्लेषणात्मक व्याख्या करना कठिन है। इसलिये अरैखिक विकास मॉडल के कुछ जैविक व्याख्यात्मक प्राचलों को आँकड़ों की श्रृंखला में शामिल करना आवश्यक है। इसके अलावा अरैखिक विकास मॉडल के प्राचलों और आर्थिक रूप से महत्वपूर्ण जैसे दूध का उत्पादन, आयु वध के समय वजन, दूध में पोषण तत्वों की स्थिति या मांस में प्राप्त पोषण तत्व के प्राचलों में संबंध की जानकारी प्राप्त कर इन्हें रोजगार की दिशा में ले जाया जा सकता है। इसलिये वर्तमान अध्ययन में बूटस्ट्रेप तकनीक के द्वारा विकास वक्र प्राचलों तथा आनुवंशिक प्राचलों के अध्ययन पर जोर दिया गया है।

सामग्री और विधि

वर्तमान अध्ययन के लिये प्रयोग किये गये आँकड़ों का एक संक्षिप्त विवरण, मॉडल का विवरण, चयनित मॉडल के प्राचलों का सचित्र प्रदर्शन किया गया। बूटस्ट्रेप तकनीक का उपयोग कर आनुवंशिक प्राचलों के गुणों का अध्ययन करने के लिये विधि दी गई है।

आँकड़ों का विवरण

इस अध्ययन के लिये वर्ष 2005, सी आई आर जी मकदूम उत्तर प्रदेश के फार्म से 142 बकरियों के शरीर के भार के आँकड़ों पर विचार किया गया।

बकरियों के शरीर के भार (किग्रा) 12 महीनों के उपलब्ध है। चार अरैखिक विकास मॉडल्स पर विचार किया गया जो इस प्रकार से हैं:

$$1. \text{ लॉजिस्टिक मॉडल } W(t) = \frac{a}{1 + be^{-t}} + e$$



2. गॉम्पर्टज मॉडल $W(t) = a \exp(-b \exp(-ct))$
3. वान-बर्टलैनफी $W(t) = a[1 - b \exp(-ct)]^b$ मॉडल
4. वीबुल $W(t) = a - (a - b) \exp[-(ct)^d]$ मॉडल
जहाँ

$W(t) = t$ आयु पर शरीर का भार

a = परिपक्व भार

b = स्थिर गुणांक

c = परिपक्व दर,

d = अतिरिक्त प्राचल

प्राचलों का आकलन एसएस संस्करण 9.1 पैकेज के द्वारा किया गया।

मॉडल चयन का मापदंड

सबसे अच्छा मॉडल चुनने के लिये निम्नलिखित मापदण्डों का उपयोग किया गया है।

- i) सबसे कम आरएमएसई (वर्ग माध्य मूल त्रुटि)
- ii) उच्चतम निर्धारण गुणांक $R^2(\%)$
- iii) निरपेक्ष न्यूनीकरण अनुपात (ARR%) (आईडूरन, 2008)

आनुवंशिक प्राचलों का आकलन

हाफ सिब मॉडल

एक सायर का एक सेट कई डैमो के साथ मैटिंग करता है और यादृच्छिक चुना जाता है। प्रत्येक मैटिंग से एक बच्चा पैदा होता है।

हाफ सिब मॉडल $Y_{ij} = \mu + s_i + e_{ij}$ है।

जहाँ μ एक सामान्य माध्य है, s_i सायर का प्रभाव है। तथा e_{ij} सायर समूहों के अंतर्गत अलग गुणों के लिये अनियन्त्रित वातावरण और आनुवंशिक विचलन है। सभी प्रभाव शून्य माध्य और निम्न प्रसरण के साथ यादृच्छिक और स्वतंत्र हैं:

$$E(s_i^2) = \sigma_s^2 \text{ और } E(e_{ij}^2) = \sigma_e^2$$

आनुवंशिकता

संकीर्ण अर्थ में यह कुल फीनोटाइपिक प्रसरण

के अनुपात को व्यक्त करता है। जो कि जेनिक को प्रमाणित (जीन का औसत प्रभाव) गुणों की नियंत्रित करता है (जैन, 1992)

गाणितीय रूप से

$$h^2 = \frac{\sigma_A^2}{\sigma_P^2}$$

जहाँ

σ_A^2 त्रयोज्य आनुवंशिकीय प्रसरण

σ_P^2 त्रफीनोटाइपिक प्रसरण हैं

एक प्रसरण का विश्लेषण तब तक किया जाता है जब तक फीनोटाइपिक प्रसरण प्रेक्षक घटकों में विभाजित हो जाए, जो कि सायर और सायर घटकों के अंतर्गत बीच में है। अगले पद में, संगत प्रेक्षण घटकों के साथ अपेक्षित घटक का वर्ग माध्य को बराबर कर σ_s^2 (सायर प्रसरण घटक) तथा σ_e^2 (प्रसरण घटक त्रुटि) का आकलन किया गया।

$$\hat{h}^2 = \frac{4 \sigma_s^2}{\sigma_s^2 + \sigma_e^2}$$

आनुवंशिकता आकलन के अनुमानित प्रतिदर्श प्रसरण (फाल्कोनर के द्वारा दिये गये, 1996) है।

$$\hat{\sigma}_{h^2}^2 = 32 \frac{\hat{h}^2}{T} \text{ (Half sib)}$$

जहाँ $T = nN, n$ और N परिवारों की संख्या प्रति व्यक्तियों की संख्या और क्रमशः परिवारों की संख्या।

आनुवंशिकता सह-संबंध

यह योज्य आनुवंशिक कारकों के दो गुणों के बीच सह-संबंध दो गुणों x और y के बीच आनुवंशिक सह-संबंध का आकलन है।

$$\hat{r}_A = \frac{\hat{\sigma}_{sxy}}{\sqrt{\hat{\sigma}_{sx}^2 \hat{\sigma}_{sy}^2}}$$

वर्तमान अध्ययन में विकास वक्र प्राचलों की आनुवंशिक प्राचलों (आनुवंशिकता तथा आनुवंशिक सह-संबंध) की मानक त्रुटि के आकलन के लिए



बूटस्ट्रेप तकनीक (ईफ्रान 1979) का प्रयोग किया गया।

प्रसरण तथा सहप्रसरण घटकों का आकलन

प्रसरण तथा सहप्रसरण घटकों का आकलन अधिकतम संभावना आकलन विधि तथा प्रतिबन्धित अधिकतम संभावना विधि से प्राप्त किया गया था क्योंकि अध्ययन के अंतर्गत जिन आँकड़ों पर विचार किया गया था वह असन्तुलित थे। आकलन की विधि पर यहाँ चर्चा नहीं की गई है। क्योंकि सभी चारों विधियों का प्रयोग कर प्रसरण घटकों के कलन की सुविधा सांख्यिकीय पैकेज जैसे SAS, SPSS के द्वारा प्रदान की गई है। सह प्रसरण घटकों के आकलन के लिये हम केम्पथोर्ने (1982) द्वारा दी गई विधि का प्रयोग करते हैं। माना हमारे पास y और z दो प्रेक्षण हैं। प्रत्येक अलग तथा समान बनाने के लिये हम स्पष्ट रूप से $y+z$ के संयुक्त प्रेक्षण पर विचार कर सकते हैं। इस संयुक्त प्रेक्षण के लिए प्रसरण के समान विश्लेषण को प्रत्येक अलग बनाने के लिये प्रत्येक y तथा z प्रारम्भिक प्रेक्षण प्रत्येक के लिये ठीक वर्गों का माध्य की अपेक्षा समान फार्मूला के लिये होगा, उन $y+z$ के लिये प्रसरण के घटक y और z के बजाय हम जानते हैं कि

$$V(y+z) = V(y) + V(z) + 2 \text{Cov}(y, z)$$

इसलिये हम लिख सकते हैं:

$$\sigma_{h(y+z)}^2 = \sigma_{hy}^2 + \sigma_{hz}^2 + 2\sigma_{hyz}$$

जहाँ

$\sigma_{h(y+z)}^2$ विविधता के स्रोत के कारण $y+z$ के लिए प्रचरण का घटक है।

σ_{hy}^2 भिन्नता स्रोत के स्रोत के लिए y के लिए विचरण का घटक है

σ_{hz}^2 भिन्नता के स्रोत के कारण z के लिए विचरण का घटक है।

σ_{hyz} भिन्नता के स्रोत के कारण y और z के लिए विचलन का घटक है। अब σ_{hyz} अनुमान के रूप में प्राप्त किया जाएगा

$$\sigma_{hyz} = \frac{1}{2} (\sigma_{h(y+z)}^2 - \sigma_{hy}^2 - \sigma_{hz}^2)$$

आनुवंशिकी प्राचल के बूटस्ट्रेप अनुमान प्राप्त करने के लिए मॉन्टे कार्लो एल्गोरिथ्म :

विभिन्न प्रकार के आयामों की संख्या में, जिसमें बूटस्ट्रेप प्राचल, विचरण, मानक त्रुटि और विश्वास अंतराल के बूटस्ट्रेप अनुमान प्राप्त करने के लिए प्राप्त किया जा सकता है। वर्तमान जाँच में नमूना, प्रक्रिया डेटा से सीधे सायर परिवार के स्तर से पुनः नमूना पर आधारित होता है। इस प्रकार से सिब परिवार मॉडल के मामले में सायर मास्टर नमूना से पुनः नमूना सीधे आधार पर पुनः प्रतिदर्शित होते रहते थे। तथा बूटस्ट्रेप इस्तेमाल करके एल्गोरिथ्म, इस प्रकार से वर्णित किया जा सकता है।

1. पूरे डेटा सामग्री को सिब परिवारों के लिए, इसी की कक्षाओं में विभाजित करते हैं। तथा F बहुभिन्नरूपी अनुभवजन्य वितरण के लिए प्रस्तुत होता है। जहाँ पर प्रत्येक परिवार के दृश्यमान $1/K$ प्रयोग किया जाता है।
2. K -वर्गों से प्रतिस्थापन के साथ एक यादृच्छिक नमूना बनाना।
3. तब महोदय परिवार में सभी संतानों का चयन करें। तथा परिवार में बच्चों का जन्मजात परिवारों के बच्चों में प्रतिस्थापन के साथ क्रमशः चुनना।
4. इस अध्ययन में अंतर्गत मॉडल के अनुसार डेटा का विश्लेषण करें।
5. अनुमान मापदंडों r और h^2 के रूप में बाधित के रूप में अनुमान लगाना।
6. 2 पर वापिस जाएँ और डेटा से एक नमूना लें और वही प्रक्रिया N बार दोहराएँ। ताकि हम N अनुमान प्राप्त कर सकें।

परिणाम और चर्चा

चार प्रकार के नामित मॉडल लॉजिस्टिक, वॉन-बर्टलैन्फी, गॉम्पर्टज और वीबुल मॉडल हैं। ये 142 वकरियों के शरीर के वजन के आँकड़ों के लिए



उपयुक्त है। वीबुल मॉडल डेटा को अच्छी तरह से फिट नहीं करता। अध्ययन के लिए इस मॉडल को नहीं लिया गया। गॉम्पर्टज, लॉजिस्टिक और वॉन-बर्टलैनफी पर मॉडल फिट किये गये। इस प्रकार से फिट मॉडल के प्राचलों का अनुमान कर मानक त्रुटि के साथ तालिका 1 में दिये गये हैं। निरपेक्ष लघु अनुपात अनुभव, R^2 RMSE, ARR का अध्ययन करनेके लिये बूटस्ट्रेप तकनीक का प्रयोग किया गया। इस प्रकार के समक्ष मापदंडों के सांख्यिकीय गुणों का अध्ययन करने के लिये 1000 नमूनों का सैम्पल लिया गया। 1000 जानवरों की संख्या जो पशुओं के तीन मॉडल में अच्छी तरह फिट होते हैं और 1000 बूटस्ट्रेप नमूने बनाये गये। चयन संतानों के स्तर पर किया गया। जो सभी तीन मॉडल्स, अर्थात् गाम्पर्टज, लॉजिस्टिक और वॉन-बर्टलैनफी प्रत्येक बूटस्ट्रेप नमूनों के सभी जानवरों में किया गया। प्रत्येक बूटस्ट्रेप नमूनों की साथ-साथ विकास मानकों और निरपेक्ष लघु अनुपात अनुमान किया गया। 110 आकार के एक बूटस्ट्रेप नमूने के लिए जिन पर तीनों माडल्स फिट होते हैं। इस संबंधित मापदंड की सामग्री विकास के लिए सभी 1000 औसतीय और उनके मानक विचलन R मापदंड के मूल्यों के औसत, अर्थात् ऊपर वर्णित मॉडल के लिए पता चलता है। बूटस्ट्रेप नमूने के परिणाम के साथ ही मास्टर नमूने के परिणाम तालिका में सारणीबद्ध हैं। तालिका न. 2 से यह देखा गया है कि बूटस्ट्रेप अनुमान मूल्य मानक त्रुटि मास्टर नमूना की तुलना में कम हैं।

R^2 मूल्यों के अनुसार, सभी मॉडल्स के R^2 मूल्य 99% से अधिक हैं। लेकिन वॉन-बर्टलैनफी मॉडल के मानक मूल्यांकन अन्य मॉडल्स की तुलना में कम हैं। अन्य मॉडल की तुलना में वॉन-बर्टलैनफी मॉडल बकरी विकास डेटा के लिए सबसे अच्छा फिट मॉडल है।

एसीआर के मान के मुताबिक, सभी मॉडल के आर का मूल्य 200 से 500 के लिए बूटस्ट्रेप नमूने का आकार बदलने में, अभी भी वॉन-बर्टलैनफी के नमूनों का आकार विकास के पैटर्न का वर्णन करने के लिए मॉडल चयन के मापदंडों को प्रभावित नहीं करता है।

वृद्धि वक्र मापदंडों के आनुवंशिकी मानकों के आकलन और उनके मानक त्रुटियां:

कार्यप्रणाली के द्वारा आँकड़ा हाफ सिब विधि द्वारा आनुवंशिकता और वृद्धि वक्र मापदंडों के आनुवंशिकी सहसंबंध के सांख्यिकीय गुण का अध्ययन करने के बूटस्ट्रेप नमूना पद्धति का उपयोग कर उत्पन्न / चयन के वृद्धि नमूना को फिट किया गया।

आनुवंशिकता का आकलन

वृद्धि वक्र मापदंडों के आनुवंशिक के बूटस्ट्रेप अनुमान लगाने के लिए नमूने के प्रथम चयन मास्टर नमूना चयनित नमूने में शामिल करने के लिए प्रत्येक प्रजनक के सभी संततियों पर विचार से प्रतिस्थापन के साथ सायर में प्रजनक स्तर पर किया गया। यदि वृद्धि शिशु जानवरों की कुल संख्या 82 है, जो प्रत्येक बूटस्ट्रेप नमूने की चयनित सायर के सभी संततियों के शरीर के भार के बराबर है। वह वॉन-बर्टलैनफी मॉडल के लिए फिट हैं यह सबसे अच्छा मॉडल विकास नमूना का वर्णन करता है। वृद्धि वक्र मापदंडों के आनुवंशिक के साथ आकलन एनोवा द्वारा किया गया। मास्टर नमूना भी अपने मानक त्रुटियों के साथ-साथ विकास मानकों के आनुवंशिकता के अनुमान प्राप्त करने का विश्लेषण करता है तथा तालिका न. 5 में आनुवंशिकता के बूटस्ट्रेप अनुमानों का वितरण तीन वृद्धि वक्र मापदंडों का अनुमान सामान्य नहीं था। तीन विकास मापदंडों की विरासत के वितरण भी पाये गये। बूटस्ट्रेप अनुमान में पूर्वान्ह तीन मापदंडों और पक्षपात के लिए अलग था। इस प्रकार से सकारात्मक मतलब यह है कि मापदंडों के बूटस्ट्रेप अनुमान में मास्टर नमूना के अनुमानों की तुलना में अधिक मूल्य है। यह परिणाम है कि आनुवंशिकता के मानक त्रुटि के सामान्य अनुमान प्राचलक 'ख' को जो मानक त्रुटि के सामान्य परिणाम होते हैं। तथा आनुवंशिकता के मानक त्रुटि के सामान्य अनुमान से बूटस्ट्रेपिंग में आनुवंशिकता अनुमान के निचले मानक त्रुटि होती है। तथा इसके अलावा बूटस्ट्रेप अनुमान से कम था। तथा यह भी देखा गया है कि बूटस्ट्रेप अनुमान के उच्च मानक त्रुटि के कारण नमूनाकरण अस्थिरता के कारण हो सकता



है। तथा जानवरों की संख्या घट तथा बढ़ सकती है। अर्थात् छोटे जानवरों की संख्या प्रमुख होती है।

तीन विकास मापदंडों पर एन्द्रापी शरीर के वजन की आनुवंशिकता अनुमान स्थिर होती है। जो प्रारम्भिक शरीर के वजन और परिवक्कता की दर से संबंधित होती है। तथा तालिका न. 5 के अनुसार मध्यम दर्जे से अधिक होती है। उच्च आनुवंशिकता असामयिक शरीर के वजन का मूल्य पाया गया है। एक ही परिणाम का साहित्य नोर्थकट और विल्सन, 1993 में पाया जाता है। इसलिए इस विशेषता के आधार पर चयन प्रभावी होना चाहिए/और इस प्राचलक के रूप में एसिमोटिक शरीर के वजन आनुवंशिकी रूप से चयन के माध्यम से बदल सकते हैं। प्राचलक भी उच्च आनुवंशिकता होती है। इसलिए यह प्रारम्भिक शरीर के वजन को बदलने के लिए पर्याप्त होती है। क्योंकि इस प्राचलक जानवर के आर शरीर के वजन से संबंधित है। जो कि चयन उद्देश्य से प्रयोग किया जा सकता है।

निष्कर्ष

विकास मापदंडों के विकास की संभावना के समान और विकास वक्र मापदंडों के बीच आनुवंशिक संबंध प्रारम्भिक चरण में सर्वश्रेष्ठ इंडीविजुअल चयन में मदद करते हैं। गैर निष्पादित पशुओं को छोड़कर समय और धन भी बचाएगा। यह प्रजनक को प्रजनन के उद्देश्य के लिये प्रारम्भिक चरण में सर्वश्रेष्ठ इंडीविजुअल का चयन करने में मदद करेगा। सामान्य तौर पर मास्टर नमूना बेहतर परिणाम प्रदान करता है। जबकि छोटे नमूना आकार के मामले में, आनुवंशिकता और आनुवंशिक संबंध और उनके मानक त्रुटियों के सटीक आकलन के लिये बूटस्ट्रेप विधि के लिये जा सकते हैं। बूटस्ट्रेप तकनीकी पक्षपाती प्रकृति को स्पष्ट रूप से तालिका से देखा गया है। यह निष्कर्ष निकाला जा सकता है कि जैसे नमूना आकार बढ़ता है, बूटस्ट्रेप नमूने के प्राचल अनुमान नमूना से बन्द हो जाता है जो कि पूर्वाग्रह का कम करना दिखाता है।

तालिका 1: फिट माडल्स के प्राचल

प्राचल	लॉजिस्टिक		गॉम्पर्टज		वानबर्टलैनफी	
	आकलन	मानक त्रुटि	आकलन	मानक त्रुटि	आकलन	मानक त्रुटि
अल्फा	32.5393	6.3961	42.9655	14.1101	37.622	6.1991
बीटा	3.7563	0.7149	1.8836	0.2706	2.0996	0.2572
गामा	0.1716	0.0320	0.0862	0.0274	0.1258	0.0457

तालिका 2 : 1000 मानक बूटस्ट्रेप नमूने और एस डी के साथ सैम्पल अनुमान के आधार पर उनके मानक विचलन के साथ संयोजित मानदंड की गुणवत्ता

	R ² (m)	RMSE (m)	ARR (m)	R ² (SD)	RMSE (SD)	ARR (SD)
मास्टर सैम्पल						
लॉजिस्टिक	0.9866	0.832	45.515	0.000159	0.036	3.456
वानबर्टलैनफी	0.9897	0.776	31.14	0.000155	0.027	3.296
गॉम्पर्टज	0.9896	0.793	35.987	0.000212	0.036	3.367



	R^2 (m)	RMSE (m)	ARR (m)	R^2 (SD)	RMSE (SD)	ARR (SD)
बूटस्ट्रेप सैम्पल						
लॉजिस्टिक	0.9976	0.853	47.502	0.00013	0.026	3.298
वानबर्टलैनफी	0.9979	0.785	32.234	0.00012	0.025	3.258
गॉम्पर्टज	0.9979	0.801	36.977	0.00012	0.025	3.280

तालिका 3: 200 मानक बूटस्ट्रेप नमूने आधार पर उनके मानक विचलन के साथ फिटिड मानदंड की गुणवत्ता

	R^2 (m)	RMSE (m)	ARR (m)	R^2 (SD)	RMSE (SD)	ARR (SD)
लॉजिस्टिक	0.9976	0.854	47.067	0.0001	0.027	3.091
वानबर्टलैनफी	0.9979	0.787	31.846	0.0001	0.026	2.996
गॉम्पर्टज	0.9979	0.802	36.541	0.0001	0.026	3.041

तालिका 4: 500 मानक बूटस्ट्रेप नमूने आधार पर उनके मानक विचलन के साथ फिटिड मानदंड की गुणवत्ता

	R^2 (m)	RMSE (m)	ARR (m)	R^2 (SD)	RMSE (SD)	ARR (SD)
मास्टर सैम्पल						
लॉजिस्टिक	0.986	0.832	45.515	0.00015	0.036	3.456
वानबर्टलैनफी	0.989	0.776	31.14	0.00015	0.027	3.296
गॉम्पर्टज	0.989	0.793	35.987	0.00021	0.036	3.367
बूटस्ट्रेप सैम्पल						
लॉजिस्टिक	0.997	0.855	47.515	0.00013	0.026	3.258
वानबर्टलैनफी	0.997	0.787	2.206	0.00013	0.026	3.183
गॉम्पर्टज	0.997	0.803	36.932	0.00013	0.026	3.257

तालिका 5: 200 विकास वक्र मापदंडों के महत्व और उनके नमूना उपायों के आधार पर मास्टर नमूना के साथ-साथ 200 बूटस्ट्रेप नमूना के संबंध में

	आनुवंशिकता (a)	आनुवंशिकता (b)	आनुवंशिकता (c)
मास्टर सैम्पल			
आकलन	0.957	0.774	0.429
मानक त्रुटि	0.491	0.481	0.450
बूटस्ट्रेप सैम्पल			

	आनुवंशिकता (a)	आनुवंशिकता (b)	आनुवंशिकता (c)
माध्य	0.977	0.834	0.636
मानक त्रुटि	0.526	0.315	0.561
माध्यिका	1.090	0.968	0.705
स्क्यूनेस	0.709	0.633	1.108
प्रसरण	0.277	0.099	0.315
कुरटोसिस	-0.070	0.396	2.128

तालिका 6: ग्रोथ वक्र मापदण्डों के बीच आनुवंशिक सहसंबंधों का अनुमान और मास्टर सैम्पल के साथ-साथ 200 वृद्धि नमूनों पर आधारित उनके सांख्यिकीय उपाय

	आनुवंशिक सहसंबंध (ab)	आनुवंशिक सहसंबंध (ac)	आनुवंशिक सहसंबंध (bc)
मास्टर सैम्पल			
आकलन	0.921	-0.318	-0.154
मानक त्रुटि	0.059	0.461	0.557
बूटस्ट्रेप सैम्पल			
माध्य	0.888	-0.313	-0.065
मानक त्रुटि	0.095	0.546	0.491
माध्यिका	0.901	-0.444	-0.084
स्क्यूनेस	-2.008	1.861	0.530
प्रसरण	0.009	0.298	0.241
कुरटोसिस	7.192	5.333	1.773

संदर्भ

1. आस्तवीन, ए.एच. (1990). यूज ऑफ बूटस्ट्रेप फॉर एस्टीमेशन ऑफ स्टैन्डर्ड डैविएशन ऐण्ड कॉन्फिडेन्स इंटरवल ऑफ जेनेटिक ऐण्ड कॉन्फिडेन्स कम्पोनेन्ट बायोमैट्रिकल जर्नल, 32: 515-527.
2. ब्राउन, जे.ई., ब्राउन, सी.जे. एवं बुटसए जर., डब्लू.टी. (1972). ए. डिस्कशन ऑफ दि जेनेटिक एसपेक्टस ऑफ वेट ऐण्ड रेट ऑफ मैच्योरिंग इन हेरेफोर्ड ऐण्ड एन्गुसकैटल. जे. एनिम. एस. सी. आई. 34: 525-536.
3. ब्राउन, जे.ई., फिटज वुग, एच. ए. एवं कार्टशाइट, टी.सी. (1972). ए. कम्पैरीजन ऑफ नॉनलीनियर माडल्स फॉर डिस्क्रीबिंग क्सेटेज रिलेशनशिप इन कैटल. जर्नल ऑफ एनिमल साइंस 42: 810-818.
4. डीनिस, आर.एस.के. (1982). रिलेशनशिप अमंग दि ग्रोल कर्व पैरामीटर ऐण्ड प्रोडक्टिविटी ट्रेटस इनग्रीफ काड, पी.एच.डी केसिस, कोलाराडो स्टेट यूनीवर्सिटी कोलाराडो.
5. एफ्रान, बी. (1979). बूटस्ट्रेप मैथडस : ऐनदर लुक एट जैकनाइफ, एनन स्टैटिस्टि. 7(1) : 1-26.
6. एफ्रान, बी. (1982). दि जैकनाइफ, दि बूटस्ट्रेप



- ऐण्ड अदर रीसैम्पलिंग प्लान्स: सी.बी.एम.एस.-एन. एस.एफ. रीजनल कानफेरेंस सीरी इन एप्लाइड मैथमैटिक्स, मोनाग्राफ 38, फिलाडेल्फिया : सियाम.
7. आईडुयन, ई, कुकुक्, एम., काराकुस, के. एवं आजडीमीर, टी. (2008). न्यू अप्रोचेज टू डिटेर्मिनेशन ऑफ द बेस्ट नॉनलिनीयर फक्शन डिस्क्राइबिंग ग्रोथ एट अर्ली फेसेज ऑफ क्रिवेरिक ऐण्ड मोरकारमन ब्रीड जर्नल ऑफ एनिमल ऐण्ड वेटेनरी एडवांसेज 7(7): 799-804.
 8. फाल्कोनर, डी.एस. (1996). इन्ट्रोडक्सन टू म्वान्टीटेटिव जेनेटिक फोर्थ एडीशन, लॉगमन ग्रुप लिमिटेड.
 9. ग्वेज, डी.पी. एवं बिजवाटर डब्लू. सी.जी. (2002). एनालाइसिस ऑफ ग्रोथ कर्वस फॉर वूडी पेरिनियल स्पीसीज, पीनस टेडा एल., थ्योरिटिकल एप्लाइड जेनेटिक्स, 105: 526-531.
 10. जैन, जे.पी. (1992). स्टैटिस्टिकल टैक्नीक्स इन क्वान्टीटेटिव जेनेटिक्स, हिन्दुस्तान पब्लिशिंग कोर्पोरेशन, दिल्ली.
 11. जयशंकर, जे. (1992). यूज ऑफ बूटस्ट्रेप टैक्नीक पर वैरिएन्स एस्टीमेशन ऑफ हेरीटेबिलिटी एस्टीमेटर्स एम अनपब्लिशड एम.एस.सी. थीसिस सबमिटेड टू इण्डियन एग्रिकल्चरल रिसर्च इन्स्टीट्यूट, नई दिल्ली-12, इण्डिया.
 12. जयशंकर, जे. (1996). बूटस्ट्रेप मेकड फॉर स्टडिंग दि डिस्ट्रीब्यूशनल प्रॉपर्टीज ऑफ एस्टीमेटर्स ऑफ वेरीऐबिलिटी, एन अनपब्लिशड पी.एच.डी. थीसिस सबमिटेड टू इण्डियन एग्रिकल्चरल रिसर्च इन्स्टीट्यूट, नई दिल्ली-12, इण्डिया.
 13. कैंचमैन, एस.डी., बाकर, आर.एफ. एवं गिआनोला, डी. (1988). फिनोटाइपिक ऐण्ड जेनेटिक वैरिएबिलिटी ऑफ एस्टीमेटेड ग्रोथ कर्व पैरामीटर्स इन मिन्स. थ्योरिटिकल एप्लाइड जेनेटिक्स, 76: 148-156.
 14. कैम्पफोर्न, ओ. (1988). एन इन्ट्रोडक्डक्शन टू जेनेटिक स्टैटिस्टिक्स, सैकेन्ड इण्डियन रीप्रिन्ट एडीशन, दि लोवा स्टेट यूनीवर्सिटी प्रैस, यू. एस. ए.
 15. कोर, ए., वास्पीनार, ई., काशका, एस. एवं केसकिन एस. (2006). द डिटरमिनेशन ऑफ ग्रोथ इनअकैसी (व्हाइट गोट) फीमेल किडस वाई वैरियस ग्रोथ फीमेल किडस वाई वैरियस ग्रोथ मॉडल्स, सीचेक जे. एनिम. साइस 51(3): 110-116.
 16. लाम्बे, एन.आर., नवजस, ई.ए., सिमम, जी. एवं बगेर आई (2006). ए जेनेटिक इन्वेस्टीगेशन ऑफ वैरियस ग्रोथ मॉडल्स ऑफ लम्बास ऑफ टू कान्ट्रासटिंग ब्रीडस. जर्नल ऑफ एनिमल साइस 84: 2642-2654.
 17. नार्थकट, एस. आई एवं विल्सन, डी.ई. (1993). जेनेटिक पैरामीटर एस्टीमेट्स ऐण्ड एक्सपैक्टेड प्रोगनी डिफरेंसेज फॉर मैच्योर साइज इन एंगुश कैटल जे. एनिम साइस 73(5): 545-548.
 18. सिंह, एन. ओकेन्द्रो एवं वाही, एस. डी. (2003). इफैक्ट ऑफ सैम्पल साइज ऐण्ड स्ट्रक्चर ऑन दि बूटस्ट्रेप एस्टीमेट ऑफ वैरिएन्स ऑफ हेरीटेबिलिटी, इण्ड. जे. एनिक साइस 73(5): 545-548.
 19. सिंह, एन. ओकेन्द्रो, राव, ए. आर., वाही, एस. डी. एवं सिंह वी.पी. (2006). रॉबस्टनेस ऑफ बूटस्ट्रेप एस्टीमेट ऑफ वैरिएन्स ऑफ हेरीटेबिलिटी अँ मास्टर सैम्पलस इन हाफ सिब एनालीसिस. इण्ड. जर्नल एनिम. जेन. ब्रीड., 27(1,2): 6-11.
 20. वाही, एस. डी., लाल चंद एवं भाटिया, वी. के. (2004). ए ग्रोथ पेटर्न स्टडी इन कोसेज ऐण्ड प्योर इण्डियन ब्रीड ऑफ गोटस इण्डियन जर्नल ऑफ एनिमल साइस 74(9): 955-958.

आभार : लेखकगण टंकण के लिए श्रीमती सुदेश अरोड़ा, निजी सहायक के आभारी हैं।

कृषि में अरफिमा मॉडल और उसके अनुप्रयोग

दिपांकर मित्र, रंजित कुमार पॉल, लाल मोहन भर, ए. के. पॉल,

हिमाद्रि शेखर राय, सविता वधवा एवं ए. आर. उदगाता

सार

इस अध्ययन का उद्देश्य स्वप्रतिगामी विभिन्नताएँ समेकित चक्रीय औसत (ARFIMA) मॉडल के सिद्धांतों एवं अनुप्रयोग की समीक्षा करना है। अध्ययन के लिए भारत में चावल के दैनिक अधिकतम, न्यूनतम और औसत थोक मूल्यों के आँकड़ों का प्रयोग किया गया। यह देखा गया कि स्थिर श्रृंखला के लिए भी मूल्य आँकड़ों में दीर्घ मेमोरी का महत्वपूर्ण स्थान है। तदनुसार आँकड़ों के लिए अरफिमा (ARFIMA) मॉडल का आकलन किया गया एवं प्रत्येक श्रृंखला के लिए पूर्वानुमान की गणना की गई। अंत में, सर्वोत्तम फिटिड अरफिमा (ARFIMA) और एरिमा (ARIMA) मॉडल का तुलनात्मक अध्ययन आर.एम.ए.पी.ई. (RMAPE), आर.एम.एस.ई. (RMSE) एवं एम.ए.ई. (MAE) के आधार पर किया गया।

परिचय

मूल्य पूर्वानुमान चक्रीय बाजार और उतार-चढ़ाव प्रतियोगिता में विश्लेषणात्मक प्रवृत्ति है। सभी व्यापारिक निर्णय एवं सरकारी नीतियाँ मूल्यों के पूर्वानुमान पर निर्भर करती हैं। अतः पूर्वानुमानों का सटीक होना महत्वपूर्ण है। गत कुछ दशकों में बॉक्स-जैनकिन्स एरिमा (ARIMA) मॉडल (बॉक्स इत्यादि, 2007) का प्रयोग पूर्वानुमानों के लिए किया गया। यद्यपि पारम्परिक स्थिरता अरमा (ARMA) प्रक्रिया प्रायः इस काल-श्रृंखला आँकड़ों में दीर्घ डिग्री उपस्थितिता को नहीं पकड़ सकी तथा अस्थिरता की वर्गमूल इकाई अथवा I(1) प्रक्रिया के कुछ निर्देश गुणात्मक गुणों के लिए वित्तीय आँकड़ों पर थी। मॉडल काल-श्रृंखला आँकड़ा प्रस्तुत करने पर दीर्घ रेंज अधिकारिकता तथा दीर्घ निर्भरता अरफिमा मॉडल (ग्रेंजर और जॉयक्स,

1980) के अनुसार प्रयोग किया गया है। पॉल (2014) और (पॉल इत्यादि, 2014) द्वारा कृषि नगों के मूल्यों के पूर्वानुमान के लिए अरफिमा मॉडल प्रयोग किया गया। परिवर्तनशीलता विवरण और पूर्वानुमान के लिए इस मॉडल ने उत्तम प्रदर्शन किया।

भारत सफेद चावल एवं ब्राऊन चावल का उत्पादक है। भारत विश्व के चावल का 20% उत्पादक है। चावल एक श्रेष्ठ फसल है और भारत में सबसे अधिक क्षेत्र में चावल की खेती होती है। चावल की भारत के समस्त भागों में खेती की जाती है। इसलिए यहाँ चावल के दामों के पूर्वानुमान के लिए उचित सांख्यिकीय मॉडल्स की आवश्यकता है। वर्तमान अध्ययन में, भारत के समस्त भागों में चावल के अधिकतम, न्यूनतम और औसत मूल्य के पूर्वानुमान के लिए अरफिमा मॉडल का आकलन किया गया तथा इसकी तुलना एरिमा मॉडल के साथ की गई।

विधियाँ

(1) लघु एवं दीर्घ मेमोरी प्रक्रिया : एक प्रतिबंधित काल-श्रृंखला प्रक्रिया $\{(Y_t)\}$ ही एक लघु मेमोरी प्रक्रिया कहलाती है, यदि $\sum_{k=0}^{\infty} |\rho_k| < \infty$ जहाँ ρ_k K लैंग के लिए काल-श्रृंखला का स्वसंबंध फलन है। सबसे बड़ी स्थिरता स्वसंबंध फलन की और परिवर्तित (ARMA) काल-श्रृंखला प्रक्रिया बहुत तेजी से एक गुणाकीय दर से नगण्य होती जा रही है इसलिए $\rho_k = |m|^k$ $|m| < 1$ ।

दीर्घ मेमोरी प्रक्रिया के लिए स्वसंबंध फलन अधिक निम्न स्तर पर होता है (परवल दर), जो $\rho_k = Ck^{-d}$ के साथ घटित होता है। जब K अनंत तक बढ़ता है जहाँ C एक स्थिरांक है और d एक दीर्घ मेमोरी प्रचालक है।



(2) एरिमा मॉडल : एरिमा मॉडल अधिकतर एक परिवर्तित और स्थिर लघु मेमोरी प्रक्रिया में प्रयोग किया जाता है। यह रेखीय एकचर काल-श्रृंखला मॉडल है जो कि एक काल-श्रृंखला प्रक्रिया $\{y_t\}, t=1,2,\dots,n$ को प्राचलकों को $\varphi(B)(1-B)^d y_t = \theta(B) u_t$ के माध्यम से स्पष्ट करता है। जहाँ y_t और u_t क्रमशः t समय पर मूल मूल्य और याच्छधिक त्रुटि हैं। $\varphi(B)$ और $\theta(B)$ यूनिट सर्कल के बाहर वर्गमूल के साथ p और q क्रम के B लैग के क्रमशः ए.आर. (AR) और एम.ए. (MA) बहुपद हैं। याच्छधिक त्रुटियाँ, u_t स्वतंत्र रूप से माध्य शून्य और प्रसारण σ^2 के साथ बंटन किये गये हैं। समान्य रूप से, ARIMA मॉडल जैसे ARIMA (p, d, q) के रूप में प्रदर्शित किया जाता है, जहाँ p, d और q स्वप्रतिगामी के क्रम को समेकित (विभिन्नता) और चक्रीय औसत को दर्शाते हैं। बॉक्स-जैनकिन्स की पद्धति के अनुसार ARIMA मॉडलिंग के तीन चरण मॉडल की पहचान, प्राचलक आकलन और निदान की जाँच हैं।

(3) अरफिमा मॉडल : दीर्घ मेमोरी की उपस्थिति में काल-श्रृंखला मॉडलिंग के लिए ARFIMA मॉडल का प्रयोग किया गया है। उदाहरण के लिए एक स्वप्रतिगामी चक्रीय औसत प्रक्रिया समेकित क्रम के साथ अरफिमा (p, d, q) निम्न है :

$$\varphi(B) y_t = (1-B)^{-d} \theta(B) u_t$$

जहाँ, $u_t, B, \varphi(B)$ और $\theta(B)$ पूर्व में भी प्रदर्शित किये गये हैं और d दीर्घ मेमोरी प्रचालक है। $d=0$ के लिए प्रक्रिया स्थिर है तथा $-0.5 < d < 0.5$ के लिए प्रक्रिया प्रत्यावर्ती और स्थिर है। $d \in (0, 1/2)$ के लिए प्रक्रिया दीर्घ मेमोरी है। दीर्घ मेमोरी प्राचलकों के आकलन के लिए अधिकतम लाईक्लीहुड आकलक (MLE) (बीरेन, 1995) और (लोकल वीटेल, (एल. डबल्यू) आकलन (रोबिंसन, 1994), जी.पी.एच. (ग्रेवेक एवं पोर्टर-हुडाक, 1983), वेवलेट पद्धति (जैनसन, 1999) इत्यादि विधियाँ प्रयोग की जाती हैं।

परिणाम एवं चर्चा

वर्तमान अध्ययन के लिए भारत में, चावल के अधिकतम, न्यूनतम और औसत मूल्य जनवरी, 2013

से फरवरी, 2017 के दौरान उपभेक्ता कार्यान्वयन निदेशालय (Consumeraffairs.nic.in) से लिए गए। कुल 941 आँकड़ों का प्रेक्षण किया गया। पहले 891 आँकड़ों का उपयोग मॉडल वैधता उद्देश्य के लिए, अन्तिम 50 आँकड़ों का उपयोग मॉडल के सत्यापन के लिए किया गया।

ऑगमेंटेड डिफे-फुलर (ADF) और फिलिप्स पैरोन (PP) टेस्ट से पता चला कि 5% महत्ता के स्तर पर यूनिट रूट टेस्ट की शून्य परिकल्पना इकाई सार्थक है जो दर्शाता है कि श्रृंखला स्थिर है। आँकड़ों के सहसंबंधित फलन (ACF) और आंशिक सहसंबंधित फलन (PACF) का भी अध्ययन किया गया और यह भी देखा गया कि सहसंबंध और आंशिक सहसंबंध दूरी लैग पर सार्थक (200 लैग के बाद भी) हैं। इससे यह स्पष्ट होता है कि मूल्य आँकड़ों में दीर्घ मेमोरी है। तालिका न. 1 के परिणामों में दीर्घ मेमोरी की उपस्थिति देखने के लिए टेस्ट किया गया और तालिका न. 1 दर्शाती है कि सभी श्रृंखलाओं में महत्त्वपूर्ण दीर्घ मेमोरी प्रतिमान है। आकलित (Z) सांख्यिकी का मान 5% महत्ता पर 1.96 से अधिक है।

तालिका 1 अरहर दाल मूल्य आँकड़ों की दीर्घ मेमोरी प्राचलक आकलन

	अधिकतम	न्यूनतम	मॉडल
डी	0.476	0.485	0.456
मा.त्रुटि	0.140	0.056	0.057
जेड	3.400	8.670	8.000

अरफिमा मॉडल दीर्घ मेमोरी प्रतिमान के लिए श्रृंखलाओं में अरफिमा मॉडल का आकलन किया गया और इसी प्रकार एरिमा मॉडल का आकलन किया गया। न्यूनतम एकेकी इंफॉर्मेशन क्रायटीरियन (AIC) के आधार पर अरफिमा मॉडल को उत्तम मॉडल चुना गया। यह देखा गया कि अरफिमा (ARFIMA) और एरिमा (ARIMA) मॉडल के प्राचल अधिकतम, न्यूनतम और औसत मूल्यों के लिए 5% महत्ता पर सार्थक हैं।

अंत के 50 प्रेक्षणों के लिए ($h=50$) एरिमा (ARIMA) मॉडल और अरफिमा (ARFIMA) मॉडल



द्वारा मूल्यों के लिए एक-स्टेप आगे के पूर्वानुमान की गणना की गई। आकलित मॉडल के सत्यापन के लिए आर.एम.ए.पी.ई. (RMAPE), आर.एम.एस.ई. (RMSE) एवं एम.ए.ई. (MAE) की गणना की गई। पूर्वानुमान का मूल्यांकन पाँच चक्रीय विण्डो (10-स्टेप 20-स्टेप, 30-स्टेप 40-स्टेप और 50-स्टेप आगे) के लिए

किया गया। प्रत्येक पूर्वानुमान के लिए भारत में चावल के अधिकतम, न्यूनतम और औसत मूल्यों के लिए आकलित आर.एम.ए.पी.ई. (RMAPE), आर.एम.एस.ई. (RMSE) एवं एम.ए.ई. (MAE) के मानों को क्रमशः तालिका न. 2, तालिका न. 3 एवं तालिका न. 4 में दर्शाया गया।

तालिका 2 अधिकतम मूल्यों के लिए पूर्वानुमान मूल्यांकन

मूल्यांकन विधियाँ	पूर्वानुमान क्षैतिज (h)					औसत
	10	20	30	40	50	
अरफिमा						
आर.एम.ए.पी.ई.	2.613	2.770	2.718	2.367	2.342	2.562
आर.एम.एस.ई.	133.638	137.829	146.956	129.632	111.909	131.993
एम.ए.ई.	115.973	122.867	121.856	117.793	104.278	116.553
एरिमा						
आर.एम.ए.पी.ई.	4.110	3.860	2.705	3.383	3.042	3.420
आर.एम.एस.ई.	190.976	196.535	173.203	172.485	154.523	177.544
एम.ए.ई.	181.340	169.132	164.108	150.409	133.936	159.785

तालिका 3 न्यूनतम मूल्यों के लिए पूर्वानुमान मूल्यांकन

मूल्यांकन विधियाँ	पूर्वानुमान क्षैतिज (h)					औसत
	10	20	30	40	50	
अरफिमा						
आर.एम.ए.पी.ई.	0.240	2.960	3.472	2.647	3.042	2.472
आर.एम.एस.ई.	4.718	84.301	60.348	65.310	59.960	54.927
एम.ए.ई.	7.000	51.718	56.820	44.914	51.460	42.382
एरिमा						
आर.एम.ए.पी.ई.	0.700	3.353	5.327	3.663	4.158	3.440
आर.एम.एस.ई.	15.462	109.610	89.937	79.389	82.434	75.366
एम.ए.ई.	11.202	60.317	87.266	61.991	70.615	58.278



तालिका 4 औसत मूल्यों के लिए पूर्वानुमान मूल्यांकन

मूल्यांकन विधियाँ	पूर्वानुमान क्षैतिज (h)					औसत
	10	20	30	40	50	
अरफिमा						
आर.एम.ए.पी.ई.	8.023	7.028	8.325	4.644	4.695	6.543
आर.एम.एस.ई.	216.00	188.596	229.644	155.298	130.136	183.935
एम.ए.ई.	181.313	167.693	195.733	116.173	120.225	156.227
एरिमा						
आर.एम.ए.पी.ई.	7.216	8.123	9.402	5.826	5.442	7.202
आर.एम.एस.ई.	218.000	250.375	248.886	187.500	144.266	209.805
एम.ए.ई.	165.884	198.742	218.633	144.559	138.874	173.338

निष्कर्ष दीर्घ मेमोरी परीक्षण के सार्थक परिणाम मूल्य श्रृंखला में दीर्घ मेमोरी के अस्तित्व को दर्शाते हैं। एरिमा (ARIMA) और अरफिमा (ARFIMA) मॉडल प्रत्येक मूल्य श्रृंखलाओं के लिए आकलित किए गए और न्यूनतम AIC मूल्यों के आधार पर उत्तम मॉडल चुना गया। न्यूनतम RMAPE, RMSE और MAE मान ARFIMA मॉडल को सभी विंडों के लिए उत्तम सिद्ध करते हैं। पूर्वानुमान के लिए एरिमा (ARIMA) मॉडल और अरफिमा (ARFIMA) मॉडल का तुलनात्मक अध्ययन किया और RMAPE, RMSE और MAE के आधार पर यह देखा गया कि अरफिमा (ARFIMA) मॉडल का प्रदर्शन एरिमा (ARIMA) मॉडल की अपेक्षा उत्तम है।

संदर्भ

1. बॉक्स, जी.ई.पी., जैनकिन्स, जी.एम. और रेनसल, पी.सी. (2007). टाईम-सीरीज़ एनालिसिस: फोरकास्टिंग एण्ड कंट्रोल, 3rd एडिशन. पियर्सन एजुकेशन, इण्डिया.

2. ग्रेवेक, जे. एण्ड पोर्टर-हुडाक., एस (1983): दी एस्टीमेशन एण्ड एप्लीकेशन ऑफ लॉग-मेमोरी टाईम-सीरीज़ मॉडल्स, जर्नल ऑफ टाईम-सीरीज़ एनालिसिस, 4: 221-238.
3. ग्रेंजर, सी. डब्ल्यू. जे. एण्ड जॉयक्स, आर. (1980). एन इंट्रोडक्शन टू लॉग-मेमोरी टाईम-सीरीज़ मॉडल्स एण्ड फ्रैक्शनल डिफरेंसिंग, जनरल ऑफ टाईम-सीरीज़ एनालिसिस, 1:15-29.
4. पॉल, आर.के. (2014). फोरकास्टिंग होलसेल प्राइस ऑफ पिजन पी यूजिंग लॉग-मेमोरी टाईम-सीरीज़ मॉडल्स, एग्रिकल्चरल इकोनॉमिक्स रिसर्च रिव्यू, 27(2): 167-176.
5. पॉल, आर.के., गुरुंग, बी. एण्ड पॉल, ए.के. (2014). मॉडलिंग एण्ड फोरकास्टिंग ऑफ रिटेल प्राइस ऑफ अरहर दाल इन करनाल, हरियाणा, दि इंडियन जर्नल ऑफ एग्रिकल्चरल साइंस, 85(1): 69-72.

विविधता के दो स्रोतों के अंतर्गत पशु परीक्षणों के लिए प्रवृत्ति मुक्त अभिकल्पनाएँ

अर्पण भौमिक, सीमा जग्गी, एल्दो वरगीस, सुनील कुमार यादव, मोहम्मद हारुन,
अनिंदिता दत्ता एवं उदयवीर सिंह

सार

किसी भी वैज्ञानिक जांच की प्रयोगात्मक सामग्री में विविधता या तो एक अथवा अधिक स्रोतों के कारण मौजूद रह सकती है। प्रयोगात्मक सामग्री में विविधता जहां पर दो स्रोतों के कारण होती हैं, उन परिस्थितियों में रो-कॉलम अभिकल्पना का उपयोग बेहतर होता है। इस लेख में व्यवस्थित प्रवृत्ति को शामिल करके रो-कॉलम मॉडल को परिभाषित किया गया है। रो-कॉलम अभिकल्पना के लिए प्रवृत्ति मुक्त अवस्था प्राप्त की जा चुकी है। इसके अलावा, प्रवृत्ति मुक्त स्तम्भ अभिकल्पना के लिए उदाहरण भी मौजूद लेख में दिये गए हैं।

प्रस्तावना

किसी भी वैज्ञानिक जांच में प्रयोग और रेखांकन वैध प्रयोग के आधार पर प्राप्त निष्कर्ष दोहरी विशेषताएँ वाले होते हैं। क्रम परिशुद्धता के स्वीकार्य स्तर के साथ परीक्षण अभिकल्पना का प्रयोग करने के दौरान किसी भी प्रयोग के लिये बहुत से कारकों (उदाहरण के लिये प्रयोगों के उद्देश्य, प्रयोगात्मक सामग्री की उपलब्धता, प्रयोग की लागत) पर सावधानी पूर्वक विचार करना चाहिये। किसी भी वैज्ञानिक जांच का एक प्रयोग की अभिकल्पना करना एक आवश्यक घटक है और इसलिये ऐसा करने के लिये दिये गये दृष्टिकोण आवश्यक है। प्रयोगात्मक सामग्री एक महत्वपूर्ण समस्या है, अतः प्रयोगात्मक सामग्री के ब्लाकिंग के माध्यम से वैज्ञानिक प्रयोगों की परिकल्पनाओं के दौरान यह ध्यान रखा जाय, कि समग्र प्रयोगात्मक सामग्री के भीतर की बजाय एक ब्लाक के भीतर की इकाइयाँ एक जैसी हों। विविधता या तो एक स्रोतों अथवा एक से अधि

क स्रोतों के कारण मौजूद रह सकती है। पशु प्रयोगों में यह स्थितियाँ काफी प्रचलित हैं। हमारे सामने कुछ परिस्थितियाँ ऐसी आ सकती हैं जहां पर दो स्रोतों के कारण प्रयोगात्मक सामग्री में विविधता का प्रमाण मिलता है। निम्न परिस्थितियों पर विचार करें।

प्रयोगात्मक स्थिति

विभिन्न चारा (उपचार) के प्रभाव की तुलना के उद्देश्य के साथ एक पशु प्रयोग में अलग-अलग गायों के दूध निकालने के विभिन्न प्रायोगिक इकाइयों को लें। मान लीजिये गाय की नस्ल और उम्र दो कारक हैं, जोकि उपचार के अलावा दो स्रोतों की विविधता के अनुरूप हैं। इस प्रकार दोनों कारक वास्तव में नियन्त्रित कारक होते हैं और यह गायों की नस्लों और उम्र के कारण भिन्नता को समाप्त करने का इरादा है। अतः यहां स्रोतों के भिन्नता को समाप्त करने के लिए ऐसी स्थिति में विविधता के दो स्रोतों के अंतर्गत अभिकल्पना का उपयोग सबसे बेहतर होता है।

पशु प्रयोगों में ज्ञात स्रोत से अलग बहुत सारी परिस्थितियाँ सामने खड़ी हो जाती हैं जैसे प्रतिक्रिया भी अस्थायी प्रभाव पर निर्भर हो सकती है। यानि प्रयोगात्मक सामग्री में एक या एक से अधिक व्यवस्थित प्रवृत्ति की उपस्थिति में प्रयोगात्मक इकाइयाँ प्रभावित हो सकती हैं। उपरोक्त प्रायोगिक परिस्थिति के लिये कोई भी व्यवस्थित प्रवृत्ति घटक को पहचान सकता है जो दूध उत्पादन को प्रभावित कर सकता है। यदि दूध उत्पादन से सम्बन्धित वास्तविकता पर ध्यान दें तो यह पहले से ही अच्छी तरह से पता होता है कि एक पशु के दूध उत्पादन की प्रगति दूध स्तनपान की अवधि में कम होती जायेगी। उपरोक्त उल्लेखित प्रवृत्ति



घटक के रूप में शामिल करने की वजह से किया प्रयोगात्मक उत्पादन अधिक सीमित संसाधनों के भीतर सही (प्री-साइज्ड) हो जायेगा।

एक तरीके से प्रायोगात्मक सामग्री में प्रवृत्तियों की उपस्थिति का पता लगाने के लिये उपयुक्त अभिकल्पना पर विचार करना है जो प्रवृत्ति प्रभाव के लिये ओर्थोगोनल होता है। ऐसी अभिकल्पना को प्रवृत्ति मुक्त डिजाइन कहते हैं। (ब्रैडली और येट, 1980) प्रवृत्ति घटकों के साथ ब्लाक अभिकल्पना के मामले में यह माना जाता है कि ब्लाक प्रवृत्ति μ^k डिग्री के ओर्थोगोनल बहुपद का प्रतिनिधित्व किया जा सकता है। जहाँ पर k ब्लाक का साईज है। राष्ट्रीय और अन्तरराष्ट्रीय स्तर पर व्यवस्थित प्रवृत्ति घटक को शामिल कर ब्लाक अभिकल्पना सेटअप के तहत काफी कार्य हो चुका है [उदाहरण के लिये : ब्रैडली और येट 1980), येट और ब्रैडली (1983), जैकाडस एट एल (1997), मजूमदार और मारटिन (2002), लाल, एट एल. (2007), भौमिक इत्यादि (2015) आदि।

इस लेख में, हमने विविधता के दो स्रोतों के अंतर्गत व्यवस्थित प्रवृत्ति को समिलित करके मॉडल को परिभाषित किया है। विविधता के दो स्रोतों के अंतर्गत एक अभिकल्पना की प्रवृत्ति मुक्त होने की शर्त को प्रकाशित भी किया गया है। इसके अलावा उपचारों की संख्या v (जो कि एक अभाज्य संख्या है) के 15 अथवा 15 से कम के लिए अभिकल्पनाओं की एक सूची भी प्रस्तुत की गई है।

प्रयोगात्मक सेट अप और मॉडल

यदि उपरोक्त प्रयोगात्मक स्थिति में दो स्रोतों की (पशु की उम्र एवं नस्ल) विविधता के बारे में वर्णित परिवर्तनशीलता की जांच करें तब विविधता के दो स्रोतों के अंतर्गत अभिकल्पना को उपयोग में लाना चाहिए। जाँचकर्ता को यह पता होता है कि एक पशु के दूध उत्पादन की प्रगति दूध स्तनपान की अवधि में कम होती जायेगी इसलिये प्रयोगों से बेहतर परिशुद्धता प्राप्त करने के लिए व्यवस्थित प्रवृत्ति के प्रभाव को मॉडल में शामिल करना चाहिये।

उपरोक्त प्रयोगात्मक सेटअप के आधार पर तय

प्रभाव निम्नलिखित मैट्रिक्स-नोटेशन में एडीटिव मॉडल माना जा सकता है।

$$Y = \mu I + \Delta' \tau + D'_1 \rho + D'_2 \chi + Z\theta + e$$

जहाँ पर Y आंकड़ों का वेक्टर है। μ एक सामान्य माध्य है। I एक वेक्टर की एकता है। Δ' एक उपचार बनाम आंकड़ों की मैट्रिक्स है। τ एक उपचार प्रभाव (जैसे चारा) का वेक्टर है। D'_1 विविधता की एक स्रोत बनाम आंकड़ों की मैट्रिक्स है जो कि नस्ल है। ρ नस्ल प्रभाव का वेक्टर है। D'_2 विविधता के दूसरे स्रोत बनाम आंकड़ों की मैट्रिक्स है जो कि उम्र है। χ उम्र प्रभाव का वेक्टर है $Z = I_v \otimes F$ जहाँ गुणांक का मैट्रिक्स है और F एक मैट्रिक्स है जिसका कॉलम ऑर्थोगोनल पॉलीनोमियल का प्रतिनिधित्व करते हैं। यहाँ e त्रुटि का वेक्टर है। जिसका माध्य शून्य और भिन्नता σ^2 है।

उपरोक्त मॉडल के आधार पर जब सभी कारकों का स्तर बराबर होता है तो उपचार के प्रभाव से सम्बन्धित आकलन की जानकारी के लिये निम्नलिखित मैट्रिक्स प्राप्त किया जा सकता है :

$$C = v \left(I - \frac{II'}{v} \right) - \frac{1}{v} \Delta Z Z' \Delta'$$

डिजाइन के लिए प्रवृत्ति मुक्त स्थिति

जब एक उचित मॉडल विनिर्देश के लिये मॉडल में व्यवस्थित प्रवृत्ति घटक शामिल होता है, ऐसी स्थिति से निपटने के लिए एक प्रभावी तरीका जो व्यवस्थित प्रवृत्ति के प्रभाव के लिये प्रतिरोधक होता है, अभिकल्पना के अनुकूल है, जहाँ पर प्रवृत्ति प्रभाव निरस्त कर दिया जाता है।

प्रवृत्ति मुक्त डिजाइन के लिये, उपचार (चारा) प्रभाव और प्रवृत्ति प्रभाव एक दूसरे के ओर्थोगोनल होते हैं और तब डिजाइन का विश्लेषण सामान्य तरीके से किया जा सकता है। अगर कोई प्रवृत्ति प्रभाव नहीं होता तो सांख्यिकीय रूप से कहा जा सकता है कि डिजाइन प्रवृत्ति मुक्त है। प्रवृत्ति घटक के साथ उपरोक्त के तहत उपचार के प्रभाव से अगर उपचार योग के वर्ग को समायोजित किया जाये और यदि प्रवृत्ति



घटक के बिना अपितु सामान्य के तहत समायोजित किया जाये तो दोनों एक जैसे होते हैं, इसलिये प्रवृत्ति मुक्त अभिकल्पना के लिये आवश्यक और पर्याप्त शर्त $\Delta Z = 0 = 0$ होती है। अतः प्रवृत्ति मुक्त डिजाइन के लिये जब सभी कारक के स्तर समान हों, जैसे (अ), तब उपचार के प्रभाव से सम्बन्धित आकलन की जानकारी के लिये मैट्रिक्स निम्नलिखित है :

$$C = v \left(I - \frac{11'}{v} \right)$$

विविधता के दो स्रोतों के अंतर्गत प्रवृत्ति मुक्त अभिकल्पनाएँ

उपचारों की संख्या $v (\leq 15)$ के लिए विविधता के दो स्रोतों के अंतर्गत प्रवृत्ति मुक्त अभिकल्पनाएँ निम्न हैं:

Design_1 (for v = 3)			
	-1	0	1
1	A 1	B 2	C 3
2	C 2	A 3	B 1
3	B 3	C 1	A 2

Design_2 (for v=5)					
	-2	-1	0	1	2
1	A 1	B 2	C 3	D 4	E 5
2	C 2	D 3	E 4	A 5	B 1
3	E 3	A 4	B 5	C 1	D 2
4	B 4	C 5	D 1	E 2	A 3
5	D 5	E 1	A 2	B 3	C 4

Design_3 (for v=7)							
	-3	-2	-1	0	1	2	3
1	A 1	B 2	C 3	D 4	E 5	F 6	G 7
2	C 2	D 3	E 4	F 5	G 6	A 7	B 1
3	E 3	F 4	G 5	A 6	B 7	C 1	D 2
4	G 4	A 5	B 6	C 7	D 1	E 2	F 3
5	B 5	C 6	D 7	E 1	F 2	G 3	A 4
6	D 6	E 7	F 1	G 2	A 3	B 4	C 5
7	F 7	G 1	A 2	B 3	C 4	D 5	E 6

Design_4 (for v=11)											
	-5	-4	-3	-2	-1	0	1	2	3	4	5
1	A 1	B 2	C 3	D 4	E 5	F 6	G 7	H 8	I 9	J 10	K 11
2	C 2	D 3	E 4	F 5	G 6	H 7	I 8	J 9	K 10	A 11	B 1
3	E 3	F 4	G 5	H 6	I 7	J 8	K 9	A 10	B 11	C 1	D 2
4	G 4	H 5	I 6	J 7	K 8	A 9	B 10	C 11	D 1	E 2	F 3
5	I 5	J 6	K 7	A 8	B 9	C 10	D 11	E 1	F 2	G 3	H 4
6	K 6	A 7	B 8	C 9	D 10	E 11	F 1	G 2	H 3	I 4	J 5
7	B 7	C 8	D 9	E 10	F 11	G 1	H 2	I 3	J 4	K 5	A 6
8	D 8	E 9	F 10	G 11	H 1	I 2	J 3	K 4	A 5	B 6	C 7
9	F 9	G 10	H 11	I 1	J 2	K 3	A 4	B 5	C 6	D 7	E 8
10	H 10	I 11	J 1	K 2	A 3	B 4	C 5	D 6	E 7	F 8	G 9
11	J 11	K 1	A 2	B 3	C 4	D 5	E 6	F 7	G 8	H 9	I 10



Design_5 (for v=13)														
	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	
1	A	1 B	2 C	3 D	4 E	5 F	6 G	7 H	8 I	9 J	10 K	11 L	12 M	13
2	C	2 D	3 E	4 F	5 G	6 H	7 I	8 J	9 K	10 L	11 M	12 A	13 B	1
3	E	3 F	4 G	5 H	6 I	7 J	8 K	9 L	10 M	11 A	12 B	13 C	1 D	2
4	G	4 H	5 I	6 J	7 K	8 L	9 M	10 A	11 B	12 C	13 D	1 E	2 F	3
5	I	5 J	6 K	7 L	8 M	9 A	10 B	11 C	12 D	13 E	1 F	2 G	3 H	4
6	K	6 L	7 M	8 A	9 B	10 C	11 D	12 E	13 F	1 G	2 H	3 I	4 J	5
7	M	7 A	8 B	9 C	10 D	11 E	12 F	13 G	1 H	2 I	3 J	4 K	5 L	6
8	B	8 C	9 D	10 E	11 F	12 G	13 H	1 I	2 J	3 K	4 L	5 M	6 A	7
9	D	9 E	10 F	11 G	12 H	13 I	1 J	2 K	3 L	4 M	5 A	6 B	7 C	8
10	F	10 G	11 H	12 I	13 J	1 K	2 L	3 M	4 A	5 B	6 C	7 D	8 E	9
11	H	11 I	12 J	13 K	1 L	2 M	3 A	4 B	5 C	6 D	7 E	8 F	9 G	10
12	J	12 K	13 L	1 M	2 A	3 B	4 C	5 D	6 E	7 F	8 G	9 H	10 I	11
13	L	13 M	1 A	2 B	3 C	4 D	5 E	6 F	7 G	8 H	9 I	10 J	11 K	12

यहां पर स्तंभ नॉन-नॉर्मलाइज्ड रेखिक प्रवृत्ति घटक को, पंक्ति विविधता के एक स्रोत को, पंक्ति-स्तंभ के कटान बिंदु के अक्षर विविधता के दूसरे स्रोत को एवं विभिन्न अंक विभिन्न उपचारों को प्रदर्शित करते हैं।

सन्दर्भ

1. भौमिक, ए., जग्गी, एस., वरगीस, सी. एवं वरगीस ई. (2014) : ट्रैंड फ्री ब्लाक डिजाइन्स बैलेन्सड फॉर इन्टरफेरेन्स इफैक्ट्स फ्रॉम नेबरिंग एक्सपेरीमेंटल यूनिट्स. जर्नल ऑफ कोम्बिनेटोरिक्स, इनफोरमेशन एण्ड सिस्टम सांइसिस. 39, (1 और 2) : 117-133।
2. भौमिक, ए., जग्गी, एस., वरगीस, सी. एवं वरगीस ई. (2015) : ट्रैंड फ्री सैकेण्ड आर्डर नेबर बैलेन्सड ब्लॉक डिजाइन्स. जर्नल ऑफ इन्डियन स्टैटिस्टिकल एसोसिएशन 53 (1-2) : 63-785।
3. ब्रैडली आर.ए. एण्ड येह, सी.एम. (1980) : ट्रैंड फ्री ब्लाक डिजाइन्स. एनल्स ऑफ स्टैटिस्टिकल 8: 883-893।
4. जैकराउस एम, मजूमदार डी. एवं शाह के आर. (1997) : ऑन दि डिटरमिनेशन ऐण्ड कन्सट्रक्शन ऑफ ऑप्टिमल ब्लॉक डिजाइन्स दि प्रेजेन्स ऑफ लीनियर ट्रैंड. जर्नल ऑफ दि अमेरिकन स्टैटिस्टिकल एसोसिएशन 92 : 375-382।
5. लाल. के. प्रसाद, आर. एवं गुप्ता, वी. के. (2007). ट्रैंड फ्री नेस्टेड बैलेन्सड इनकम्प्लीट ब्लॉक डिजाइन्स फॉर डायलल क्रॉस एक्सपेरीमेंट्स. कलकत्ता स्टैटिस्टिकल एसोसिएशन बुलिटैन 59 : 203-221।
6. मजूमदार, डी. एवं मारटिन, आर.जे., (2002) फाइंडिंग्स ऑप्टिमल डिजाइन्स इन दि प्रेजेन्स ऑफ ट्रैंड्स. जर्नल ऑफ स्टैटिस्टिकल प्लानिंग एण्ड इन्टरफेरेन्स. 106, 177-190।
7. येह, सी.एम. एवं ब्रैडली, आर. ए. (1983) : ट्रैंड फ्री ब्लॉक डिजाइन्स. अगिस्टेन्स ऐण्ड कन्सट्रक्शन रिजल्ट्स, कम्प्यूनिकेशन इन स्टैटिस्टिकल-थ्योरी ऐण्ड मैथड्स. 12 (1) : 1-24।

माइक्रोसैटेलाइट डी.एन.ए. मार्कर के प्रयोग से मवेशी नस्ल पहचान हेतु वेब सर्वर

मीर आसिफ इकबाल, सारिका, ऊषा जैन, अनिल राय एवं दिनेश कुमार

सार

घरेलू गाय (बॉस टौरस) विभिन्न गुणों के लिए जैसे – दूध की पैदावार, मांस की गुणवत्ता, ड्राफ्टेबिलिटी, रोग एवं कीटों के प्रति प्रतिरोध तथा सामाजिक एवं धार्मिक कारणों के लिए मनुष्य द्वारा चुनी गयी प्रजातियों में से एक है। नवपाषाण युग (8,000 से 10,000 साल पहले) से आज तक, मवेशी पालन की संख्या 1.5 मिलियन हो गयी है तथा आगामी वर्ष 2050 तक 2.6 बिलियन होने की सम्भावना है। संख्याओं के उच्च परिमाण, नस्ल प्रबन्धन, नस्ल उत्पाद की ट्रैसेबिलिटी, संरक्षण प्राथमिकता एवं जर्मप्लाज्म पलो/विनिमय के कारण आई.पी.आर. मुद्दों ने सटीक एवं तीव्र नस्ल पहचान के लिए महत्वपूर्ण हैं। वर्षों से नस्ल की पहचान के लिए परिभाषित नस्ल डिस्क्रिप्टर्स का प्रयोग किया गया है परन्तु ओवा, वीर्य, भ्रूण तथा प्रजनन उत्पादों में फिनोटिपिक वर्णन के अभाव के कारण मॉलीक्यूलर अप्रोच अनिवार्य है। इसके अतिरिक्त, सम्मिश्रण की डिग्री एवं गैर-वर्णक्रम पशुओं के लक्षण वर्णन के लिए मॉलीक्यूलर अप्रोच की आवश्यकता अनिवार्य है। आज तक, मॉलीक्यूलर डाटा विश्लेषण पर आधारित नस्ल पहचान विधियों की बहुत सी सीमाएँ हैं जैसे कि संदर्भ डाटा उपलब्धता की कमी तथा कम्प्यूटेशनल विशेषज्ञता की आवश्यकता। इन चुनौतियों पर काबू पाने के लिए हमने संदर्भ डाटा के रखरखाव तथा नस्ल पहचान हेतु सुविधाओं के लिए एक वेब सर्वर विकसित किया (<http://webapp.cabgrid.res.in/biscattle>)। प्रागुक्त मॉडल विकसित करने के लिए प्रयुक्त संदर्भ डाटा 8 मवेशी नस्लों और 18000 एलील डाटा उत्पन्न करने वाले 18 माइक्रोसैटेलाइट डी.एन.ए. मार्कर्स से प्राप्त किए गए। इस अध्ययन में, महत्वपूर्ण loci की पहचान एवं loci संख्या में कमी करने के

लिए विभिन्न एल्गोरिथ्म का प्रयोग किया गया। 95 प्रतिशत सटीकता के साथ मेमोरी-आधारित लर्निंग एल्गोरिथ्म का प्रयोग करते हुए 5 लोकस प्राप्त की गयीं। यह मॉडल अप्रोच एवं पद्धति, नस्ल पहचान और संरक्षण कार्यक्रम में विश्व भर में सभी घरेलू पशुओं की प्रजातियों में बहुत बड़ी भूमिका निभा सकती है। इसे सभी वनस्पतियों और जीवों के लिए भी उनकी विभिन्न किस्मों की पहचान करने एवं जर्मप्लाज्म प्रबन्धन के लिए भी मॉडल किया जा सकता है।

प्रस्तावना

घरेलू गाय (*Bos taurus*) आर्थिक एवं सांस्कृतिक रूप से विश्वभर की एक महत्वपूर्ण प्रजाति है जो समस्त मानव आबादी को पोषण उपलब्ध कराती है। मानव द्वारा विश्व भर में 800 विभिन्न मवेशी नस्लों का विभिन्न गुणों अर्थात् दूध उत्पादन, मांस गुणवत्ता, ड्राफ्ट क्षमता, रोगों एवं कीटों के प्रति प्रतिरोधिता, तथा सामाजिक एवं धार्मिक कारणों के लिए चयन किया। मवेशियों का पशुपालन घरेलू कारणों के लिए नवपाषाण युग में (8000–10000 वर्ष पूर्व) प्रारम्भ हुआ। इस समय 1.5 बिलियन से अधिक मवेशिया रिपोर्ट की गयीं हैं जो खाद्य एवं कृषि संगठन के अनुसार वर्ष 2050 तक 2.6 बिलियन तक की वृद्धि के लिए उत्तरदायी है (FAO 2007)। मॉलीक्यूलर स्तर पर नस्लों की बहु संख्या होने के कारण एवं नस्ल प्रबन्धन के लिए एक सटीक पहचान टूल की आवश्यकता है।

प्रत्येक नस्ल एक दिए गए पारिस्थितिक स्थान में भू-जलवायु के रिस्पॉस में जीनपूल अनुकूलता में विकसित जीन्स का एक अनूठा संयोजन है। पशुधन पशुओं में, प्रजनन उद्देश्यों के लिए व्यक्तिगत पहचान आवश्यक है क्योंकि उनकी क्षमताओं को अगली पीढ़ी



तक सीधे पास कर दिया जाता है। मवेशियों में पूर्व में लागू की गयी विधियों में टैटूज एवं कान के टैग का प्रयोग किया गया था जिसमें व्यक्तिगत पहचान संख्या थी, उसके बाद अधिक जानकारी वाली इलेक्ट्रॉनिक चिप्स थीं। समय के साथ, पैरेन्टेज जाँच के लिए प्रोटीन पॉलीमॉरफिज्म पर आधारित रक्त जीनोटाइपिंग परिणाम का भी अनुसरण किया गया परन्तु रक्त जीनोटाइपिंग तंत्र की परीक्षात्मक जटिलता के कारण, इसे डी. एन.ए. से प्रतिस्थापित किया गया ठीक उसी तरह जिस तरह से मानव पर फॉरेन्सिक विज्ञान में लागू किया जाता है।

आजकल, नस्लों के वर्गीकरण के लिए, नस्ल सांसाइटियों अथवा वैधानिक निकायों द्वारा घोषित सुपरिभाषित नस्ल डिस्क्रिप्टर्स का प्रयोग किया जाता है। इन फिनोटिपिक डिस्क्रिप्टर्स की कुछ सीमाएँ हैं क्योंकि वीर्य, ओवा, भ्रूण अथवा नस्ल उत्पाद की पहचान के लिए इनका उपयोग नहीं किया जा सकता। इसके अतिरिक्त, ये फिनोटिपिक डिस्क्रिप्टर्स सम्मिश्रण में नस्ल की प्रागुक्ति नहीं कर सकते। अनेक अध्ययनों में घरेलू पशु नस्ल की पहचान के लिए एस.टी.आर. मार्कर्स का प्रयोग किया गया है।

हमने यहाँ 8 मवेशी नस्लों के परीक्षण डाटा का प्रयोग करते हुए नस्ल पहचान के लिए एक आदर्श मॉडल अप्रोच प्रस्तुत की है। कम जीनोटाइपिंग लागत का लाभ पाने के लिए इस पद्धति का उपयोग अधिक संख्या की नस्लों/देशों में किया जा सकता है चूंकि संदर्भ डाटा हर बार जीनोटाइपिंग की लागत को खत्म करता है। वेब सर्वर आधारित कम्प्यूटेशन इस अप्रोच को अधिक प्रयोक्ता फ्रेंडली बनाता है।

सामग्री एवं विधियाँ

डाटा जनरेशन/उपलब्धता

संदर्भ एलीलिक डाटा जनरेंट करने के लिए, संबंधित देशी प्रजनन प्रणाली में यादृच्छिक स्तरित प्रतिचयन का प्रयोग करते हुए हमने कुल 500 नमूने एकत्रित किए। रक्त के नमूने वैक्यूटेनर में एकत्रित किए गए जिसमें ई.डी.टी.ए. (EDTA) एन्टीकॉग्यूलेंट के रूप में था। प्रत्येक नस्ल के आनुवंशिक रूप से

संबंधित नमूने एकत्रित करने के लिए उचित ध्यान रखा गया। 500 नमूनों के पैनल में डांगी (67) खिल्लर (66), निमारी (65), माल्वी (68), कन्क्रेज (56), गिर (33), गैलाओ (75), केनकाथा (70) पशु नस्ल शामिल थे। खाद्य एवं कृषि संगठन (FAO) पशु आनुवंशिकी हेतु अन्तरराष्ट्रीय सांसाइटी और पूर्व अध्ययनों द्वारा तथा सुझाए गए माइक्रोसेटेलाइट सबबप पैनल से 18 माक्रोसेटेलाइट सबबप का चयन किया गया (चौधरी, इत्यादि 2009) कैले, इत्यादि, 2010)। ये 18 loci हैं CSRM60, ILSTS005, ILSTS011, ILST006, MM12, ILSTS030, BM1824, HAUT27, BM1818, ETH152, INRA035, ETH10, INRA005, ILSTS034, CSSM663, ETH3, INRA063 एवं ILSTS033. चूंकि माक्रोसेटेलाइट मार्कर्स सह-प्रभावशाली (co-dominant) हैं अतः 18 सह-प्रभावशाली सबबप एवं 500 नमूनों के संयोजन से नस्ल सिग्नेचर का संदर्भ डाटा विकसित करने के लिए अध्ययनगत समष्टि के लिए 18000 एलिलिक डाटा जनरेंट की प्रत्याशा है। सभी नमूने की प्रत्येक लोकस के एलिलिक रिचनेस (Rt), दुर्लभ (rarefaction) विधि एवं नस्लों के बीच युग्मवार आनुवंशिक दूरी का आकलन राइट के FST (θ) के वेयर एवं कॉकरहैम (1984) आकलनों का प्रयोग करते हुए FSTAT v2.9.3 (गोडेट, 2002) से किया गया।

सांख्यिकीय अप्रोच

मवेशी नस्लों के वर्गीकरण के लिए सटीक मॉडल के निर्माण हेतु बेसियन नेटवर्क, निकटतम नेबर (IB1) एवं स्पोर्ट वेक्टर मशीन (SVM) सहित मेमोरी-बेस्ड लर्निंग जैसे विभिन्न क्लासीफायर्स एप्लाई किए गए। मॉडल विकसित करने के बाद, सटीकता, मैथ्यू का सहसंबंध गुणांक (MCC), संवेदनशीलता, विशिष्टता, सकारात्मक प्रागुक्त मान (PPV) एवं नकारात्मक प्रागुक्त मान (NPV) जैसे मूल्यांकन मापों के माध्यम से मॉडल की प्रागुक्ति की गुणवत्ता की जाँच की गयी। इस अध्ययन में, प्रागुक्ति-त्रुटि के आकल प्राप्त करने के लिए फाइव-फोल्ड क्रॉस वैलीडेशन तकनीक लागू की गयी। सटीकता, मैथ्यू का सहसंबंध गुणांक (MCC), संवेदनशीलता, विशिष्टता, सकारात्मक प्रागुक्त मान (PPV) एवं नकारात्मक प्रागुक्त मानों



(NPV) के सांख्यिकीय आकलन के माध्यम से मॉडल की प्रागुक्ति की गुणवत्ता की जाँच की गयी। मूल्यांकन मापों को निम्नानुसार परिभाषित किया गया है :

$$\text{Sensitivity} = TP / (TP + FN)$$

$$PPV = TP / (TP + FP)$$

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

$$\text{Specificity} = TN / (FP + TN)$$

$$NPV = TN / (TN + FN)$$

$$MCC = \frac{(TP * TN - FP * FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

सर्वर का क्रियान्वयन

यह वेब सर्वर (<http://webapp.cabgrid.res.in/biscattle>) Hyper Text Markup Language एवं जावा का प्रयोग करते हुए विकसित किया गया है। यह एक प्रयोक्ता-फ्रैण्डली वेब सर्वर है जो वेब सर्वर सॉफ्टवेयर अपाचे का प्रयोग करते हुए लॉच किया गया है। इसमें .csv फॉरमेट, .txt फॉरमेट अथवा सबमिशन फार्म में डाटा सीधे अपलोड करने का विकल्प है। इस सर्वर में अनेक लचीलेपन (flexibilities) जोड़े गए हैं। प्रयोक्ता को 5 लोकस, 10 लोकस तथा 18 लोकस के साथ नस्ल पहचान का विकल्प होता है। इसमें 7 टैब्स हैं अर्थात् होम, सबमिशन, एल्गोरिथ्म, ट्यूटोरियल, टीम, लिंक्स एवं एफ.ए.क्यू (FAQs)। इस सर्वर में प्रयोक्ता को आसानी से समझाने के लिए एक सरल डाटा के साथ ट्यूटोरियल की भी सुविधा है।

परिणाम एवं चर्चा

इस अध्ययन में, नस्ल पहचान के लिए कम्प्यूटेशनल जटिलताओं को न्यून करने व लोकस की संख्या में कमी करने वाले एक मॉडल वेब सर्वर को विकसित करने हेतु माइक्रोसेटेलाइट डी.एन.ए. मार्कर डाटा का उपयोग किया गया। इसे प्राप्त करने के लिए, मवेशी

के 18000 एलीलिक/माइक्रोसेटेलाइट डाटा पर तीन क्लासीफायर्स अर्थात् बेसियन नेटवर्क, मेमोरी-आधारित लर्निंग (IB1) एल्गोरिथ्म एवं सपोर्ट वेक्टर मशीन लागू किए गए। यह पाया गया कि विकसित किए गए मॉडल्स में फाइफ-फोल्ड वैधीकरण के साथ क्रमशः 95.45 प्रतिशत, 97 प्रतिशत तथा 89.65 प्रतिशत सटीकता थी। इस लिए, मेमोरी-आधारित लर्निंग (IB1) एल्गोरिथ्म का उपयोग करते हुए विकसित मॉडल को अंततः TiMBL: टी.आई.एम.बी.एल. मेमोरी-बेस्ड लर्निंग वर्जन 6.1 सॉफ्टवेयर को उपयोग करके सर्वर पर क्रियान्वित करने के लिए चुना गया। संवेदनशीलता, विशिष्टता तथा एमसीसी क्रमशः 88, 98.29 एवं 0.86 पायी गयी (तालिका 1)। क्रियान्वित एल्गोरिथ्म की नस्लवार सटीकता एवं एमसीसी क्रमशः 91 प्रतिशत से 100 प्रतिशत तथा 0.63 से 1.00 के बीच थी (तालिका 2)।

हमारे परिणाम 97 प्रतिशत सटीकता के साथ 18 विभिन्न loci सहित मेवशियों की 8 नस्लों की पहचान का समर्थन करते हैं। यह दर्शाता है कि locus की संख्या में वृद्धि से सटीकता एवं एम.सी.सी में क्रमशः 88.50 प्रतिशत से 97.00 प्रतिशत तथा 0.47 से 0.86 वृद्धि होती है (तालिका 3)। हम सिर्फ पाँच loci (CSRM60, ILSTS005, BM1824, ILSTS034, ETH3) का चयन करके 0.80 एमसीसी के साथ 95.55 प्रतिशत तक सटीकता प्राप्त कर सकते हैं। लोकस की संख्या को कम करते समय, हमने पाँच एस.टी. आर. लोकस का प्रयोग करते हुए 95.55 प्रतिशत सटीकता तथा दस लोकस के साथ 96.1 प्रतिशत सटीकता प्राप्त की। हमारी जाँच से पता चलता है कि दस लोकस से परे, लोकस जीनोटाइपिंग की कोई आवश्यकता नहीं है, अतः यह नस्ल पहचान की 50 प्रतिशत लागत की बचत है। आठ और अतिरिक्त लोकस का शामिल करने से सटीकता में केवल 10 प्रतिशत तक की वृद्धि होगी जिससे जीनोटाइपिंग की लागत में लगभग दोहरे परिणाम में वृद्धि होगी, जो वांछनीय नहीं है। हमें लोकस के पाँच शीर्ष रूचिकर मान प्राप्त हुए जो अत्यधिक विभेदित हैं (FST > 0.15, $\omega_a R_t > 8.0$)। शीर्ष 10 लोकस के मामले में, दो लोकस



(अर्थात् HAUT27 एवं INRA035) को छोड़कर शेष में समान रेंज मिली (तालिका-4)। हमने यह भी पाया कि जहाँ आनुवंशिक भिन्नता (FST मान) अपेक्षाकृत कम हैं, एलीलिक रिचनेस (Ri मान) काफी अधिक थी ताकि सम्भावित नस्ल की पहचान के लिए लोकस की सूचनात्मकता की क्षतिपूर्ति हो सके। यद्यपि, उपलब्ध बोवाइन उच्च घनत्व (HD) SNP चिप (778k) (बाई, इत्यादि, 2012) तथा न्यून घनत्व (एलडी) एसएनपी चिप (54k) (क्यूहेन इत्यादि 2013) का उपयोग नस्ल पहचान के लिए भी किया जा सकता है परन्तु फिलहाल, विश्व के अधिकांश भागों के लिए ये लागत प्रभावी करते हुए, सॉफ्टवेयर की सीमित संख्या द्वारा किया जा सकता है, उदाहरणतया मॅडल (लेंज, इत्यादि, 2013), जो पुनः सर्वर मॉड में नहीं है जिससे प्रयोक्ता-फ्रैण्डली न होने संबंधी मुद्दे शामिल होंगे।

निष्कर्ष

वर्तमान अध्ययन में घरेलू पशु नस्ल प्रागुक्ति के लिए विश्व का पहला मॉडल वेब सर्वर रिपोर्ट किया गया है। हमने 5, 10 एवं 18 लोकस के साथ 8 मवेशी नस्लों की क्रमशः 95.5 प्रतिशत 96.10 प्रतिशत तथा 97 प्रतिशत सटीकता रिपोर्ट की है। लोकस की कम संख्या से न केवल लागत में तीव्र कमी आएगी अपितु सम्मिश्रण की डिग्री के साथ मॉलीक्यूलर स्तर पर नस्ल की पहचान के लिए अधिक कम्प्यूटेशनल सुगमता भी प्रदान करेगा। यह संप्रभुता एवं जैव-पायरेसी विवाद के मामले में आईपी संरक्षण के साथ मौजूदा नस्ल एवं नई कृत्रिम वाणिज्यिक नस्लों के लिए एक अपरिहार्य टूल हो सकता है। इस वेब सर्वर का अन्य घरेलू प्रजातियों के लिए एक मॉडल के रूप में इस्तेमाल किया जा सकता है, साथ ही, पूरे विश्व भर में सभी वनस्पतियों और जीवों के जर्मप्लाज्म प्रबन्धन में भी उपयोग किया जा सकता है। यद्यपि, हमने इस मॉडल को माइक्रोसेटेलाइट डीएनए मार्कर्स पर विकसित किया है परन्तु जीनोमिक्स एवं कम्प्यूटेशनल टूल्स के लाभ प्राप्त करने के लिए एसएनपी चिप आधारित डाटा की प्राप्ति हेतु संदर्भ डाटा आधारित समान सर्वर आधारित अप्रोच आवश्यक की जा रही है।

संदर्भ

- बाई वाई, सारटर, एम एवं कवलकोली, जे (2012) – करेन्ट स्टेटस ऐण्ड पयूचर पर्सपेक्टिव फोर सीक्वेन्सिंग लाइवस्टॉक जीनोमस। ज.एनि.साई. बायोटेक्नो. 3 : 8
- चौधरी एमवी, परमार एसएनएस, जोशी सीजी, भोंग सीडी, फातिमा एस, इत्यादि (2009) – मॉलीक्यूलर कैरेक्टराइजेशन ऑफ ब्रीड्स युजिंग माइक्रोसेटेलाइट मार्कर्स। एनि. बायोडायवर, कन्सर्व. 32 : 71-76
- एफएओ (2007) – दि स्टेट ऑफ दि वर्ल्ड्स एनिमल जेनेटिक्स रिसोर्सेज फॉर फूड ऐण्ड एग्रिकल्चर। <http://www.fao.org/docrep/010/a1250e/a1250e00.htm>.
- गोडेट जे. (2002) – एफएसटीएटी, ए प्रोग्राम टू एस्टीमेट ऐण्ड टेस्ट जीन डायवर्सिटीज ऐण्ड फिक्सेशन इन्डिसेज वर्जन 2.9.3.2 उपलब्ध : <http://www2.unil.ch/popgen/softwares/fst-at.htm> (गोडेट, 1995 से अद्यतन)
- कैले डीएस, रैंक डीएन, जोशी सीजी, यादव बीआर, कौरिंगा पीजी, इत्यादि (2010) – जैनेटिक डायवर्सिटी अमंग इण्डियन गिर, दयोनी ऐण्ड कन्क्रेज कैंटल ब्रीड्स बेस्ड ऑन माइक्रोसेटेलाइट मार्कर्स। इण्डियन ज. बयाटेक्नो 9 : 126-130
- क्यूहेन एलए, कीले जेडब्ल्यू, बेनेट जीएल, मैकडोनेल्ड टीजी, स्मिथ टीपीएल, इत्यादि (2013) – प्रीडिक्टिंग ब्रीड कम्पोजीशन यूजिंग ब्रीड प्रीडिक्न्सीज ऑफ 50,000 मार्कर्स फ्रॉम दि यू.एस. भीट एनिमल रिसर्च सेन्टर 2,000 बुल प्रोजेक्ट। ज.एनि.साई. 89 : 1742-1750
- वेयर, बीएस, कुकरहेम सीसी (1984) – एस्टीमेटिंग स्टैटिस्टिक्स विद स्पेशल रिगार्ड टू सिस्टम ऑफ मैटिंग। इवोल्यूशन (एनवाई) 38 : 1358-1370
- लेंज कं, पैप, जंसी, इत्यादि (2013) – मॅडल : दि स्विस् आर्मी नाइफ ऑफ जेनेटिक एनालिसिस प्रोग्राम। बायोइन्फॉरमेटिक्स, 29(12) : 1568-70

तालिका 1 : विभिन्न क्लासीफायरों का निष्पादन

क्लासी-फायर्स	संवेदन शीलता	विशिष्टता	सटीकता	एमसीसी
बेसियन नेटवर्क	81.80	97.40	95.45	0.79
सपोर्ट वेक्टर मशीन	58.60	94.09	89.65	0.53
TiMBL-IB1 एल्गोरिथ्म	88.00	98.29	97.00	0.86

तालिका 2 : मवेशी की आठ नस्लों के लिए प्राप्त प्रागुक्त सटीकताएँ

नस्ल	संवेदन शीलता	विशिष्टता	सटीकता	एमसीसी
DAG	100.00	100.00	100.00	1.00
GAO	100.00	100.00	100.00	1.00
GIR	60.61	99.57	97.00	0.73
KAN	69.64	96.62	93.60	0.67
KEN	100.00	100.00	100.00	1.00
KHL	100.00	100.00	100.00	1.00
MAL	80.88	96.53	94.40	0.76
NIM	73.85	93.56	91.00	0.63
Wt. Avg	88.00	98.29	97.00	0.86

तालिका 3 : बढ़ती हुई लोकस के साथ संवेदनशीलता, विशिष्टता, सटीकता एवं एमसीसी की, उनके रैंक के अनुसार, औसत के लिए प्रकृति

लोकस की संख्या	संवेदन शीलता	विशिष्टता	सटीकता	एमसीसी
1	54.00	93.43	88.50	0.47
2	70.00	95.71	92.50	0.66
3	73.80	96.26	93.45	0.70

लोकस की संख्या	संवेदन शीलता	विशिष्टता	सटीकता	एमसीसी
4	77.60	96.80	94.40	0.74
5	82.20	97.46	95.55	0.80
6	81.80	97.40	95.45	0.79
7	83.00	97.57	95.75	0.81
8	83.00	97.57	95.75	0.81
9	83.00	97.57	95.75	0.81
10	84.40	97.77	96.10	0.82
11	85.40	97.91	96.35	0.83
12	85.00	97.86	96.25	0.83
13	85.20	97.89	96.30	0.83
14	86.00	98.00	96.50	0.84
15	86.40	98.06	96.60	0.84
16	87.20	98.11	96.95	0.85
17	87.80	98.26	96.95	0.86
18	88.00	98.29	97.00	0.86

तालिका 4 : शीर्ष 10 लोकस के Rt एवं FST

लोकस	Rt (Allelic richness)	FST (Genetic differentiation)
CSRM60	13.69	0.150
ILSTS005	8.20	0.270
BM1824	8.17	0.272
ILSTS034	14.83	0.151
ETH3	8.09	0.359
ILSTS030	5.36	0.150
HAUT27	8.32	0.088
ETH152	9.33	0.321
INRA035	11.66	0.093
INRA005	8.23	0.113

कृषि में कीट संख्या की गतिशीलता पर आईएनएआर (INAR) मॉडल का अनुप्रयोग

हिमाद्रि शेखर राय, रंजित कुमार पॉल, एल. एम. भर, मृन्मय राय,
प्रकाश कुमार, अचल लामा, दीपंकर मित्र एवं राजु कुमार

सार

वास्तविक संसार की घटनाओं में, कुछ स्थितियों में प्रायः ऐसा होता है, जहाँ निरंतर टिप्पणियाँ नहीं होती हैं। काल-श्रृंखला में ऐसी स्थितियाँ होती हैं। जैसे सड़क दुर्घटनाओं की संख्या, अस्पताल में भर्ती रोगियों की संख्या, फोन कॉल की संख्या एवं बेची गई उत्पादों की संख्या इत्यादि, कृषि में फसल के मौसम में उत्पन्न कीटों की संख्या, एक निश्चित प्रकार की बीमारी से संक्रमित फसलों की संख्या हो सकती है। इनमें, परिस्थितियाँ, सामान्य, स्वचलित (ए.आर.) मॉडल आदि, जहाँ गौशी वितरण का पालन करने के लिए त्रुटियाँ लागू नहीं होती हैं। इसलिए यहाँ एक आवश्यकता है कि गिनती टिप्पणियों के लिए, मॉडल विकसित तैयार करने के लिए गणना अवलोकन प्रयोग किये गये मॉडल पॉसॉन प्रतीपगमन (Poisson Regression) होता है। इस कारण चर, कर्णाक स्वचलित (आई.एन.ए.आर.) मॉडल गिनती टिप्पणियों की भविष्यवाणी के लिए प्रयोग किया जाता है, जहाँ वास्तविक डेटा सेट लागू होते हैं। तथा कीट आबादी पर पूर्वानुमान किया गया है। अंत में (आई.एन.ए.आर.) मॉडल पॉसॉन प्रतीपगमन (Poisson Regression) के साथ तुलना की जाती है और यह देखा जाता है, कि, (आई.एन.ए.आर.) मॉडल पॉसॉन प्रतीपगमन मॉडल की तुलना में अच्छा पूर्वानुमान देता है।

कुंजी शब्द: स्वचलित मॉडल, पूर्वानुमान, गॉसियन बंटन, पूर्णाक स्वचलित मूल्य मॉडल, पॉसॉन प्रतीपगमन काल-श्रृंखला

वास्तविक दुनिया की स्थिति में, हम अलग-अलग परिस्थितियों का सामना कर सकते हैं। जहाँ परिस्थितियों में टिप्पणियाँ सामान्य नहीं होती हैं। तथा समय के

अंतराल में श्रृंखला-ढांचा से सहसंबंध गिनती ही मिल सकती है। इस प्रकार से इन सभी की सहसंबंध गिनती कहलाती है। टिप्पणियों के अनुसार मॉडल विकसित करने की आवश्यकता है। और इस तरह के लिए गणीतीय काल-श्रृंखला का वर्णन आवश्यक होता है। तथा टिप्पणियों का प्रकार ऐसी काल-श्रृंखला में दिखाई देता है। जो कई वास्तविक जीवन स्थितियाँ, जैसे दुर्घटनाओं की संख्या, अस्पताल में भर्ती रोगियों की संख्या, फोन कॉल की संख्या, अपराध में होने वाली संख्या, उत्पीड़न, बेचने वाले उत्पादकों की संख्या एवं विकास इत्यादि इस तरह के पूर्णाक के लिए मॉडल और सांख्यिकीयों के निष्कर्ष का मूल्य काल-श्रृंखलाओं की टिप्पणियों को दो भागों में विभाजित किया जाता है। पॉसॉन (Poisson Regression) प्रतीपगमन मॉडल और पूर्णाक मूल्यवान ऑटोरिग्रेसिव चक्रीय औसतन (ARMA) मॉडल

आर्य इत्यादि (2015) में प्रकाशित (ARIMAX) मॉडल, मौसम चर का उपयोग करके कीट आबादी का अनुमान लगाया जाता है। जब एक प्रक्रिया में अत्यधिकता होती है, तो पूर्णाक को हम निरंतर अनुमानित करते हैं। मानक स्वचलित प्रक्रिया व्यवहारित होती है, इसलिए, मूल्य की पूर्णाक का उपयोग करने की आवश्यकता होती है। मॉडल में मार्कॉव चेन और उसके गुणों को प्राप्त करने के लिए (Cox & Miller, 1965) में प्रस्तुत किया था। लेकिन विकसित ही पैरामिट्रिक मॉडल ही विकसित किये थे। और इनकी यही सीमाएं ये थीं। कि और उनके संबंध में संरचना की शर्तें और फिर जेकब एवं लुईस ने (1978) में असतत करार दिया। तथा ऑटोरिग्रेसिव की शुरुआत की। एवं (आई.एन.ए.आर.) आवेदन करके, चक्रीय औसत (आई.एन.ए.आर.),



मॉडल का निर्माण किया। तथा ये विभिन्न तरीकों से बनाये गये थे। इस प्रकार के मॉडल को उपयुक्त बनाने के लिए कई लेखकों, अतः पूर्वानुमान मूल्यां के साथ कुछ परिणाम प्राप्त किये गये थे। पूर्णांक मूल्यांकित स्वचलित मॉडलों के लिए (आई.एन.ए. आर.) का प्रयोग करना मुख्य रूप से (आई.एन.ए.आर.) मॉडल (पतले-पन) पर आधारित थे। चालकों द्विपदीय सूक्ष्म चालक और इसकी सूक्ष्म सामान्याकरण (स्टूटेल, वॉनहार्न एवं (हरन) द्वारा पहली बार पेश किया था। हरन में सूक्ष्म काम व्यापक रूप से किये थे। INAR मॉडल विकसित करने के लिए कार्यान्वित मैकेन्जी (1985) द्विवार्षिक विकसित द्विपक्षीय गणना की टिप्पणियों के लिए स्वचलित मॉडल की संरचना अच्छी तरह व्याख्यात्मक है। तथा गिनती टिप्पणी की स्थिरता अल-जोश और अल-ओश (1978) में पहली बार पूर्णांक मूल्यवान यादृच्छिक मॉडल विकसित किया गया था। जहाँ INAR एक प्रक्रिया के रूप में जाना जाता है। और अंतराल के लिए सबसे उपयुक्त है। यह भी दिखाया गया कि वितरण गुण और सह-संबंध मॉडल की संरचना निरंतर रहती है। मूल्यवान प्रगतिशील या ए.आर. प्रक्रिया विभिन्न आकलन प्रक्रिया जैसे अधिकतम संभावना आकलन प्रक्रियाएं इत्यादि अधिकतम संभावना, आंकलन (एम.एल.ई.) सशर्त कम से कम वर्ग (सी.एल.एस.) और यूले-आकार (वाई.डब्लू) की भी पद्धति का वर्णन किया गया था। आलओश (1990) ने (आई.एन.ए.आर.) INAR (1) मॉडल को विकसित किया था। पीथ-आर्डर जो कि असततः समय के मॉडलिंग के लिए उपयोगी है। और निर्भर गणना प्रक्रिया को उन्होंने अंतर को दिखाया था। सहसंबंध में गॉसियन एवं ए.आर. (पी) प्रक्रिया के साथ मार्कोवियन संपत्ति और प्रतीपगमन स्वतंत्रता और अस्वतंत्रता को ध्यान में रखते हुए डेटा के कमी में, थाइरेगौड एट-अल (1999) विकसित बारिश की गतिशीलता को मॉडलिंग के लिए एवं विधि द्वारा नमूना एक (अपिंग-गाल्टी) बारिश गेज अलग-अलग मॉडलों के साथ लगी संख्या प्रस्तुत की गई थी। और आकलन किया गया कि मॉडल की प्रक्रिया को भी वर्णित किया गया था। करले और शर्मा (2014) के लिए जाँच की संभावना मॉडल प्रमुख कीटनाशकों

कीट संख्या की गतिशीलता को समझते हुए, चावल आलू-ओकरा फसल सिस्टम के तहत बू-एवं मैककेग (2008) में अनुमान लगाया था। कि मूल्य निर्धारण की एक श्रृंखला के लिए मॉडल चयन प्रक्रिया कि भी श्रृंखला के लिए आदमियत मॉडल आकलन संभावना प्रदर्शन करके ही किया जाता है। तथा मॉडलों के लिए एक नई पद्धति पेश की थी। तरीकों एवं अवशेषों के आधार पर चयन पूर्वानुमान प्राप्त किये गये थे। तथा विश्वास अंतराल के साथ आंकलन त्रुटि और जैसे मार्कव चैन (श्रृंखला) के आधार पर मॉडल को व्यवहारित किये थे। पायलोपोलस एण्ड कार्लिश (2008) को दूसरे प्रकार के INAR (1) (आई.एन.ए.आर.) मॉडल को दर्शाए गये थे। तथा श्रृंखलाओं की लम्बाई कुल समय पर मार्कव श्रृंखला स्वप्रतीपगमन की संरचना अरैखीय के बारे में जो चर्चा की गई थी। वह पॉसान लॉ (Poisson Law) के एक अनिश्चित मिश्रित वंटन के लिए त्रुटि की गई थी। ई-एम (EM) एल्गोरिथ्म का प्रयोग करके सशर्तः लघुमान अधिकतम के द्वारा निष्कर्ष, बूटस्टेप, प्रचालक पूर्णांक मूल्य पूर्वानुमान (IP) की पहुँच के लिए विकसित और निश्चित सांख्यिकीय प्रयोग करके संयोजित मॉडल की कार्यशीलता को प्रयोग के लिए चर्चा की गई थी।

इनसिसो (Enciso) मोरा (Mora) इत्यादि (2009) में विकसित की गई INAR(1) प्रक्रियाएँ मॉडल में चरों की व्याख्याओं के समक्ष आँकड़ा मॉडल क्रमांक के लिए सबसे अच्छा होता है। एक प्रयाप्त एम.सी.एम. सी. (MCMC) एल्गोरिथ्म विश्लेषण का चयनित आधार और दोनों व्याख्याओं चरों की महत्ता संयोग और मॉडल विश्लेषण के लिए सरंचरित की गई थी।

क्रमांक आँकड़ा विश्लेषण के लिए, पूर्णांक मूल्यकृत काल-श्रृंखला द्वारा बहुत से प्रायोगिक निराकरण प्रयोगों के लिए, विशेष रूप से लघु बारंबारता विस्तृत आँकड़ा प्रयुक्त होता है। किंतु उसके निरंतर रूप से लगभग परिणाम अच्छा नहीं होता है। (फ्रीलैंड एवं एम.सी. कैब) (2003) के अनुसार, यह INARMA एवं (ARMA) पूर्णांक मूल्यांकित से संरचना की गई है। तथा नील एवं राव (2005) के अनुसार पूर्णांक परिणाम संरचित किये गये हैं।



इस पेपर में, हमने INAR(1) मॉडल को परिचालित किया है। इस प्रकार से, हमारी प्रथम प्राथमिकता कृषि में गतिशीलता के द्वारा कीटनाशक जनसंख्या को पूर्वानुमानित करती है। जहाँ द्विवास्तविक आँकड़ा सेट INAR(1) के द्वारा विश्लेषण किया गया है। जो कि पॉसोन प्रतिगमन के साथ मॉडल (INAR)(1) के द्वारा मॉडलों की तुलना की गई है। INAR(1) मॉडल फ्रेमवर्क काल-श्रृंखला विस्तृत परीक्षणों के अध्ययन के लिए यह देखा गया है कि, इनको अपनी कसौटी पर जाँचा-परखा गया है।

सामग्री एवं विधियाँ

मान लीजिए कि $\{x_t\} t=0,1,2,\dots,n$ क्रमांक आँकड़ा काल-श्रृंखला एम अनिश्चित सीमा के साथ $(0,\dots,n)$ एक क्रमांक का है। जहाँ $n \in N = \{1,2,\dots\}$ और एक काल-श्रृंखला पर निर्भर स्वप्रतिगामी (AR) गौसीयन (Gaussian) के साथ जानी जाती है। जहाँ Auto-regressive (AR) की प्रक्रिया प्राप्त होती है। तथा सीमांत बटन द्विपदीय बटन, उदा. $B(n,p)$ जहाँ $PE(0,1)$ द्विपदीय (Binomial) A.R. मॉडल कहलाता है। जो कि प्रथम वार एम.सी. कैंजी द्वारा प्रस्तावित था। द्विपदीय (AR) के मामले में वीव एवं किम (2013) में वर्णित सहसंबंध संरचना क्रमिक काल-श्रृंखला के साथ प्रथम आदेश के अनुसार क्रमवित है। एसद्रॉपिक बटन सशर्त लघु-वर्ग आकलकों के साथ द्विपदीय AR(1) मॉडल के साथ भी चर्चा की गई है।

INAR(1) मॉडल : पूर्णांक मूल्यांकित स्वप्रतिगामी मॉडल के लिए प्रथम आदेश पर एम.सी. कैंजी के द्वारा अल-ओश एवं अलजैद (1987) ने स्वतंत्र पूर्णांक निर्भर क्रमांक प्रक्रिया श्रेणीबद्ध पूर्वानुमान और मॉडल के लिए प्रचालकों को प्रचालित किया है। यदि $\{x_t\}$ विस्तृत गणात्मक पूर्णांक (अनकारत्मक पूर्णांक) मूल्यांकित स्टॉकैस्टिक प्रक्रिया जब INAR(1) मॉडल अंतराल स्टॉकैस्टिक समी. दी गई है।

$$X_t = \alpha * X_{t-1} + \varepsilon_t$$

जहाँ $\alpha > 0$ तथा $\{*\}$ एक सोचनीय आप्रेशन को प्रदर्शित करता है और $\{\varepsilon_t\}$ एक (अ-तृणात्मक) पूर्णांक मूल्यांकित श्रेणियाँ है। तथा μ_ε एक σ_ε^2 माध्य के साथ

चर और चर तथा INAR(1) मॉडल व्याख्या दर्शाई गयी निम्न प्रकार से है

$$X_t = \underbrace{\alpha * X_{t-1}}_{\text{समय पर जनसंख्या}} + \underbrace{\varepsilon_t}_{\text{इमीग्रेशन (प्रवासी)}}$$

सोचनीय आप्रेशन का प्रयोग करके X_t एक चयनित व्यक्तिगत संख्या की तरह होता है। इसलिए INAR(1) प्रक्रिया श्रृंखला पद्धति प्रक्रिया INAR(1) मॉडल के गॉल्टन-वेस्टन के समक्ष प्रक्रिया की गई है। एवं बहुत से लेखकों का INAR(1) मॉडलों को प्रयोग करके इस मॉडल के बहुत से अनुप्रयोग है। जैसे कि उदा. के लिए X_t ग्राहकों की संख्या प्रदर्शित करता है। ε_t ग्राहकों एवं नये ग्राहकों की संख्या प्रदर्शित करता है। $X_{t-1} - \alpha * X_{t-1}$ ग्राहकों की संख्याएं जो कि अंतिम समय का अंत तक जीवित नहीं रही।

जनसंख्या INAR(1) मॉडल के लिए $\{X_t\}$ दिये गये समी. में द्विपदीय थिंकर ऑपरेटर के गुणों का प्रयोग करके प्रस्तुत किया है।

$$\text{जहाँ } \mu_x = E(X) = \frac{\mu_\varepsilon}{1-\alpha} \text{ और } \sigma_x^2 = \frac{\alpha\mu_\varepsilon + \sigma_\varepsilon^2}{1-\alpha^2}$$

लेग की संख्याओं के लिए INAR(1) मॉडल रैखीय पूर्वानुमान के लिए रैखीय समी. को पबलोयोलिस और कार्लिस (2008) ने प्रस्ताव किया था जो निम्न है। जैसे कि:

$$\hat{X}_{t+1} = \alpha^t X_t + (1-\alpha) \frac{\mu_\varepsilon}{1-\alpha}$$

उपरोक्त समी. रैखीय समी. की अधिकतम आधुनिक परीक्षणों की है। जिनके माध्य (Mean) एवं प्रसारण (Variance) निम्न है। जैसे कि:-

$$E(\hat{X}_{t+1}) = \alpha^t \mu + (1-\alpha^t) \frac{\mu_\varepsilon}{1-\alpha} \text{ और } Var(\hat{X}_{t+1}) = \alpha^{2t} \sigma^2$$

उदाहरणतः इस अनुभाग में हम, कृषि में कीटनाशक जनसंख्या की पूर्वानुमान के मॉडल INAR(1) मॉडल प्रस्तावित एक अनुप्रयोग के द्वारा विश्लेषण करते हैं। तथा आँकड़ा सेट को कृषकों के क्षेत्र में विभिन्न प्रकार के मौसमी साप्ताहिक आँकड़ों से आंध्रप्रदेश के गंटूर जिले में कृषकों ने (2008 से 2012) तक के दौरान में व्हाइटफ्लाई और जेसिड की संख्या पर



ऑकडे एकत्र किये गये थे। ये ऑकडे (R-Software) के द्वारा विश्लेषित किये गये थे। इस अध्ययन में चर औसत संख्या के कीटनाशक 03 स्तरों पर (यादृच्छिक चयन) के द्वारा किया गया था तथा ऑकडों की सूची 110 मौसमी कथन के द्वारा सप्ताहों में की गई थी। तथा प्रत्येक वर्ष में, 22 मौसमी विभागों में 32 सप्ताह से 52 सप्ताहों के बीच और एक सप्ताह की सफलता वर्ष में कीटों की प्राप्ति के अनुसार टुकड़ों में रखा गया है। जैसिद जनसंख्या के मामलों में अधिकतम संख्या कीटनाशकों की दर्ज की गई थी। जैसे कि (2010-11) के 39 सप्ताहों में रखी गई थी इस प्रकार व्हाइटप्लाइ के मामलों में, अधिकतम कीटों की संख्याओं को दर्ज किया गया था जैसे कि 45 सप्ताहों को 13 कीटों की संख्याओं के बीच (2010-11) में प्रस्तुत की गई थी। यह भी पता लगाया गया था कि चरता प्रसारण गुणांक तथा 63% जैसिद जनसंख्या का प्रसारण गुणांक होता है। तथा इससे कम नहीं होता है और 125% व्हाइटप्लाइ से कम है। तथा 125% व्हाइटप्लाइ कम अनुपात होता है। जैसिद जनसंख्या मामलों में माध्य (1.52) से प्रसारण (0.93) अधिक रहता है। तथा माध्य (2.45) से अधिक रहता है। विस्तृत सांख्यिकी ऑकड़ा सेट के साथ तालिका न.1 में प्रस्तुत की गई है। काल-श्रृंखला प्लॉट के ग्राफीय प्रतिवेदन स्वप्रतिगामी फलन और आंशिक स्वसंबंध मिलन आकृति न. 1 और 2 में दर्शाई गई है। आकृति न. 2 में एक विस्तृत रिपोर्ट नीचे तथा ऊपर 01 वर्ष के दौरान दर्शाई गई है और 01 वर्ष के समय की वृद्धि मॉडल समाप्त हो जाते हैं। तथा प्रकृति में स्थाई जनसंख्या ऑकड़े भी दो कीटनाशक जनसंख्या के लिए समय सामप्ती के पश्चात भी स्वप्रतिगामी परिणाम आकृति न. 1 और आकृति न. 2 में घनात्मक प्रस्तुत किये गये हैं।

परिणाम और चर्चा: व्हाइटप्लाइ जनसंख्या के लिए INAR(1) को परीक्षणों से आकलित प्रचालकों के साथ यूल-वाल्कर विविधा प्रयोग करके अलग से जैसिद जनसंख्या के लिए $\lambda = 0.32$ एवं $\alpha = 0.80$ को द्वि ऑकड़ा सेट के परीक्षणों के साथ INAR(1) मॉडल को फिट किया जाता है। जब पॉसान प्रतिगामी मॉडल

को प्रयोग करके INAR(1) मॉडल से प्राप्त द्विऑकड़ा सेट के लिए कीटनाशक जनसंख्या का पूर्वानुमान किया गया है। तथा INAR(1) मॉडल के साथ-साथ पॉसान प्रतिगामी का तुलनात्मक अध्ययन तालिका न. 2 में प्रस्तुत होता है। दोनों कीटनाशक मॉडल की वैधता के लिए मौसमी सप्ताह 49, 50, 51, 52 और 1 को सम्मानित किया गया है। किंतु INAR(1) मॉडल का पूर्वानुमान (1.90) है। और पॉसान प्रतिगामी का पूर्वानुमान (1.27) है। तो 49जी सप्ताहों के लिए मूल पूर्वानुमान जैसिद के लिए होता है। और पॉसान प्रतिगामी पूर्वानुमान (1.14) और INAR(1) पूर्वानुमान (1.90) है। परीक्षणों की संख्या 50जी सप्ताहों के लिए बराबर होगी। पॉसान प्रतिगामी पूर्वानुमान (2.79) है और (0.90) के साथ 03 वस्तुतः परीक्षण 51 सप्ताह के लिए प्रयुक्त होती है। इसी के साथ INAR(1) पूर्वानुमान (1.90) है। व्हाइटप्लाइ के मामलों में सामान्य परीक्षण प्राप्त होते हैं। इस तुलनात्मक अध्ययन से प्राप्त होता है कि हम पॉसान प्रतिगामी पूर्वानुमान INAR(1) मॉडल काफी अच्छा होता है। हम कम सकते हैं। कि आकृति न.1 यह भी सूचित करती है कि INAR(1) मॉडल बिल्कुल संतोषजनक पाया गया है तथा वर्गमूल माध्य मानक त्रुटि (RMSE) और एकाके सूचना अधिकृत (AIC) मूल्य INAR(1) मॉडल और पॉसान प्रतिगामी मॉडल की तुलना की गई है। और ये मूल्य तालिका न. 3 में दर्शाए गये हैं। फलस्वरूप, व्हाइटप्लाइ जनसंख्या और जैसिद के लिए 420.8 और 237.62 मॉडल INAR(1) मॉडल मूल्यों के लिए INAR(1) मॉडल को (AIC) मूल्यों को सापेक्ष में किया जाता है। इसी के साथ (RMSE) मूल्य INAR(1) मॉडल (0-16) और (0-46) जैसिद और व्हाइटप्लाइ जनसंख्या के समक्ष होते हैं तथा दोनों ऑकड़ा सेट (RMSE) मूल्यों और (AIC) मूल्यों के INAR(1) मॉडल पॉसान प्रतिगामी मॉडल से कम है।

क्रमांक ऑकड़ा विश्लेषण के लिए, सबसे अधिक INAR(1) मॉडल और पॉसान प्रतिगामी द्विमॉडल के लिए प्रयोग किया जाता है। वर्तमान खोज में INAR(1) मॉडल को ही परिचयता की प्राथमिकता दी गई थी। यह भी इस पेपर में चर्चा पूर्वानुमानता के INAR(1)



तालिका 1: आँकड़ा सेट की विस्तृत सांख्यिकी

कीटों के नाम	न्यूनतम	अधिकतम	बहु.	मध्शिका	माध्य	प्रसरण	प्रसरण गुणांक
जैसिद	0	5	1	1	1.52	0.93	0.63
व्हाइटपलाई	0	13	0	1	2.54	9.38	1.25

तालिका 2 : पॉसान प्रतिगामी और INAR(1) के द्वारा भविष्यकृत परीक्षणों के विरुद्ध वस्तुतः परीक्षणों की तुलना करना

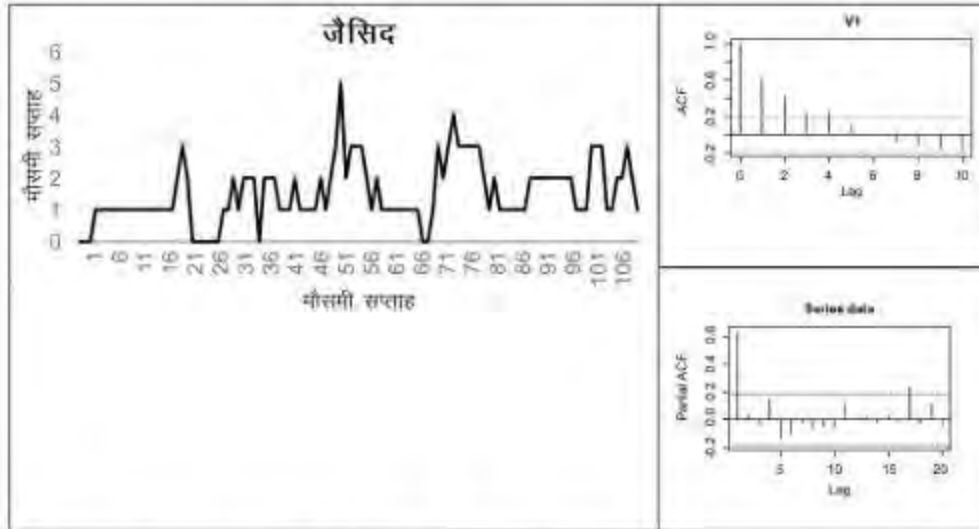
कीट	कीटों की तुलना	2013 के मौसमी सप्ताह				
		49th	50th	51th	52th	1st
जैसिद	वस्तुतः परीक्षण	2	2	3	2	1
	INAR(1) पूर्वानुमान	1.90	1.90	2.70	1.91	1.12
	पॉसान प्रतिगामी पूर्वानुमान	1.27	1.14	0.90	1.12	1.07
व्हाइटपलाई	वस्तुतः परीक्षण	3	2	3	2	2
	INAR(1) पूर्वानुमान	3.48	2.2	2.85	2.85	2.23
	पॉसान प्रतिगामी पूर्वानुमान	2.42	4.84	5.11	3.57	4.66

तालिका 3 : (RMSE) वर्गमूल माध्य त्रुटि और एकाके सूचना मॉडल (AIC) Akake Information Model INAR(1) मॉडल और पॉसान प्रतिगामी मॉडल

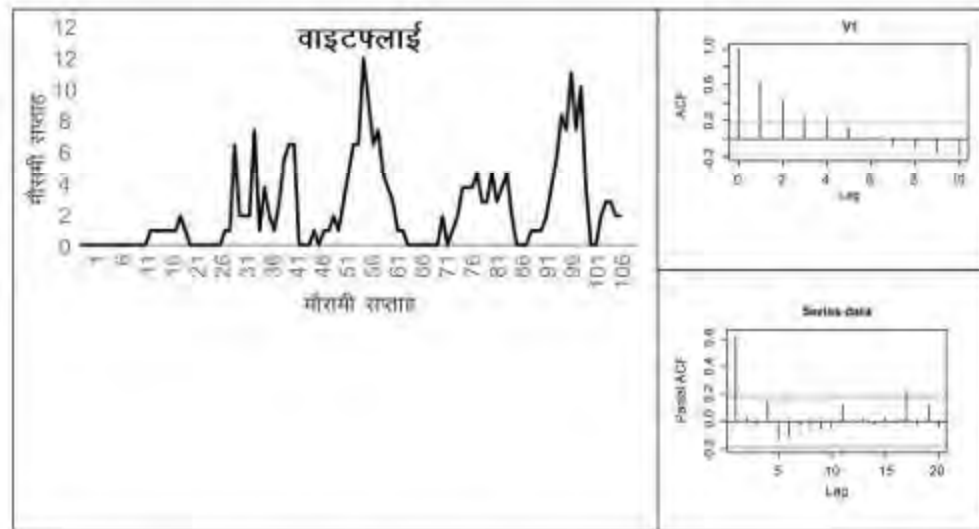
कीट (PEST)	वर्गमूल माध्य त्रुटि (RMSE)	(AIC)	
जैसिद	INAR (1)	0.16	237.62
	पॉसान प्रतिगामी	0.98	302.95
व्हाइटपलाई	INAR(1)	0.46	420.81
	पॉसान प्रतिगामी	1.31	490.68

मॉडल में वास्तविक आँकड़ा सेट को अलग-अलग किये गये थे तथा INAR(1) मॉडल के साथ-साथ पॉसान प्रतिगामी आँकड़ा सेट की तुलनात्मक अध्ययन कीटनाशक जनसंख्या के पूर्वानुमान को मालूम करने के

लिए किया गया था। विस्तृत काल-श्रृंखला परीक्षणों के INAR (1) मॉडल अन्तिम खोज एक अच्छा पूर्वानुमान किया जाता है।



आकृति 1: जैसिद जनसंख्या का कार्यफलन और आंशिक स्वसंबंध नमूना स्वसंबंध का काल श्रेणी प्लोट

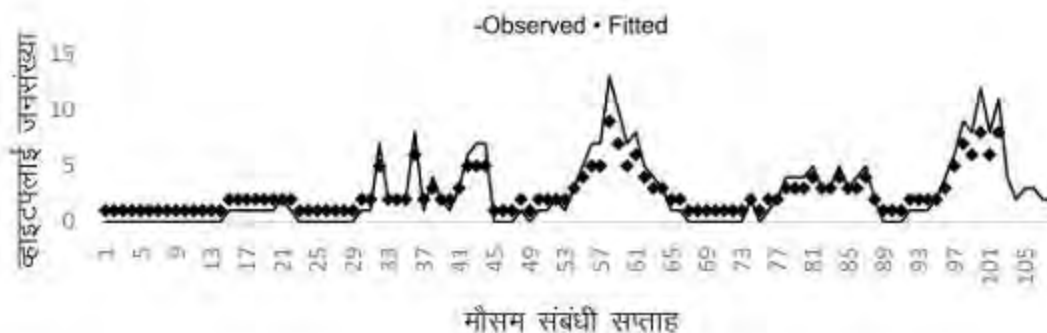


आकृति 2: द्हाइटपलाई जनसंख्या का कार्यफलन और आंशिक स्वसंबंध नमूना स्वसंबंध का काल श्रेणी प्लोट जैसिद जनसंख्या





आकृति 3: जैसीद और व्हाइटपलाई जनसंख्या का देखे जनसंख्या (observed) और परिकलित (fitted) जनसंख्या
व्हाइटपलाई जनसंख्या



संदर्भ

1. एल-आस, एम.ए. एवं अलजैद, ए.ए. (1987). फस्ट ऑर्डर इंटीगर-वेलयूड ऑटारिग्रेसिव INAR(1) प्रोसेस. ज. टाईम सीरीज अनाल., 8: 261-75.
2. अलजैद, ए.ए. और एल-आस, एम.ए. (1990). एन इंटीगर-वेलयूड पथ-आर्डर ऑटारिग्रेसिव स्ट्रक्चर INAR(1) प्रोसेस. ज. ऐप-प्रोब, 27: 314-24
3. आर्य, पी., पॉल, आर. के., कुमार, ए., सिंह, के. एन., सिवारमने, एन. एवं चौधरी, पी. (2015). प्रीडिक्टिंग पैस्ट पॉपुलेशन यूजिंग वेदर वेरियेबलस: एन अरिमेक्स टाईम-सीरीज फ्रेमवर्क. इन्ट., ज. एग्रीक. स्टेट. साई. 11(2): 381-86.
4. ब्रीनालस, के. एवं हेलोस्ट्राम, जे. (2002). ए न्यू एप्रोच टू मॉडलिंग ऐण्ड फोरकास्टिंग मंथली गेस्ट नाईटस इन होलटेल्स इंडियन, ज. फोरकास्टिंग, 18: (19.30)।
5. बू, आर. एवं मैककेब, बी. (2008)। मॉडल सेलेक्शन, एसटीमेशन ऐण्ड फोरकास्टिंग इन INAR(1) मॉडलस: ए लाइकलीहुड-बेसड मार्कोव चैन एप्रोच, इन्ट., ज. फोरकास्टिंग, 24: 151-62.
6. कोक्स, डी. आर. एवं मिलर, एच.डी. (1965)। दि थ्योरी ऑफ स्टाकास्टिक प्रोसेज, कैम्पमैन ऐण्ड हाल, सी.आर.सी., वाशिंगटन, डी.सी.
7. ड्यू, जे. जी. एवं वाई, ली. (1991)। दि इंटीजर्स वेलयूड ऑटारिग्रेसिव INAR(p) मॉडल, ज. टाईम-सीरीज अनाल, 12: 129-42.
8. एनसीसॉ-मोरा, वी., नील, पी. एवं राव टी. एस. (2009)। इंटीजर्स वेलयूड प्रोसेसज विद एक्सप्लेनेटरी वेरीएबलस, संख्या, दि इण्डि. ज. स्टैट. 71(B): 248-63.
9. जेकॉवस, पी. ए. एवं लेविस, पी. ए. डब्लू. (1978)। डिस्क्रीट टाईम-सीरीज जनरेटिड बाई मिक्सचर्स I: कोरिलेशन ऐण्ड रन्स प्रोपर्टीज: सीरीज, ज. रॉयल स्टैट. सोस., 40(B): 94-105.
10. कार्ल, एम. एवं शर्मा, एच. एल. (2014). ऑन प्रोबेबिलिटी मॉडलस फॉर डिस्क्राबिंग पॉपुलेशन डायनामिक्स ऑफ मेजर इनसेक्ट पेस्टस अंडर राइस-पॉटेटो-ओकरा क्रोपिंग सिस्टम, ज. क्रोप वीड. 10: 451-56.
11. मैककैनजिक, ई. (1988). सम सीपल माडलस फॉर डिस्क्रीट वेरिएट टाईम-सीरीज, वाटर रिसार्सेज बुल., 21: 645-50.
12. मैककैनजिक, ई. (1988). सम अरमा मॉडलस फॉर डीपेंडेंट सीक्वेंसेज ऑफ पॉसन कारंटस, अडव. एप. प्रोब., 20: 82-35.



13. नील, पी. एवं राव, टी. एस. (2005). एम.सी.एम. सी. फॉर इंटीजर्स वेल्यूड अरमा प्रोसेसज, रिसर्च रिपोर्ट, 10 प्रोवेबिलिटी ऐण्ड स्टैटिस्टिक ग्रुप स्कूल ऑफ मेथमेटिक्स, दि यूनिवर्सिटी ऑफ मानचेस्टर.
14. पवलोपुलस, एच. एवं कारलिस, डी. (2005). INAR(1) मॉडलिंग ऑफ ओवरडिसर्पस्ड काउंट सीरीज विद् एन एन्वायरमेंटल ऐप्लीकेशन. इन्वायरमेंट्रिक्स, 19: 369-93.
15. स्टूटल, एफ. डब्लू. एवं वेन हरन, के. (1979). डिस्क्रेट ऐनोलोग ऑफ सेल्फ-डिकम्पोजिटिविटी ऐण्ड स्टैबिलिटी, दी एनल्स ऑफ प्रोव. 7: 893-99.
16. थाइरेगोड, पी. कारसटनसैन, जे., मडसैन, एच. एवं अमुवजरग-नीलासन, के. (1999). इंटीजर्स वेल्यूड आटोरेग्रेसिव मॉडल्स फॉर टिपिंग बुकैट रैनफॉल मेजरमेंटस, इन्वायरमेंट्रिक्स, 10: 395-411.
17. वीब, सी. एच. एवं किम, एच. (2013). बायोनोमियल AR(1) प्रोसेसज: मूमेंटस कियूमूलेंटस, ऐण्ड एस्टीमेशन, स्टैटिस्टिक्स, 47: 494-510.

एरिमा काल श्रृंखला मॉडलिंग

रामसुब्रमणियन वी., रविन्द्र सिंह शेखावत, के. एन. सिंह एवं एस. पी. सिंह

परिचय

काल श्रृंखला समय के क्रम में प्रेक्षण की एक श्रेणी है। अधिकांशतः ये प्रेक्षण समान रूप से स्थान, असत समय अन्तराल पर एकत्रित किये जाते हैं। जब केवल एक चर है जिस पर प्रेक्षण बना रहे हैं तब हम उन्हें एकल काल श्रृंखला कहते हैं या अधिक विशेष रूप से एक विचर काल श्रृंखला कहते हैं। किसी भी काल श्रृंखला विश्लेषण/मॉडलिंग में एक बुनियादी धारणा है कि पिछले पैटर्न के कुछ पहलु भविष्य में जारी रहेंगे, इस प्रक्रिया के तहत भी, अक्सर काल श्रृंखला प्रक्रिया मुख्य चर के पिछले मान के आधार पर कल्पना की जाती है लेकिन व्याख्यात्मक चर पर नहीं जो चर प्रणाली को प्रभावित करते हैं। इसलिये यह तंत्र एक "ब्लैक बॉक्स" के रूप में कार्य करता है क्योंकि हम सिर्फ यह जान सकते हैं कि क्या घटित होगा बजाय इसके कि क्यों घटित होगा इसलिए यदि काल श्रृंखला मॉडल का उपयोग, पूर्वानुमान के प्रयोजनों के रूप में होता है तो वे विशेष रूप से अल्पावधि में लागू होंगे। यहाँ यह मौन रूप से माना जाता है कि अतीत के बारे में जानकारी संख्यात्मक आंकड़ों के रूप में उपलब्ध है। आदर्श रूप से, काल श्रृंखला विश्लेषण/मॉडलिंग के लिए कम से कम 50 प्रेक्षण आवश्यक हैं, जैसाकि बॉक्स और जेनकिंस द्वारा प्रस्तावित किया गया था जो काल श्रृंखला मॉडलिंग में अग्रणी थे।

जहाँ तक कृषि में काल श्रृंखला मॉडलिंग की उपयोगिता का संबंध है, सांख्यिकीय पूर्वानुमान मॉडलिंग के क्षेत्र में इस के अनुपयोग का मुश्किल से ही किसी जोर की आवश्यकता हो। समय पर पूर्वानुमानों का अभाव, जैसेकि, कृषि उत्पादन के लिए विशेष रूप से अल्पकालीन पूर्वानुमान अक्सर योजनाकारों के लिए एक बड़ी बाधा साबित हुआ है। विभिन्न सांख्यिकीय दृष्टिकोणों जैसे कि प्रतिगमन, काल श्रृंखला और स्टॉकास्टिक

दृष्टिकोण, फसल पूर्वानुमानों के लिए प्रचलित हैं। हर दृष्टिकोण के अपने फायदे और सीमाएँ हैं। कुछ स्थितियों में काल श्रृंखला मॉडल के उचित रहे हैं। उन्हें पूर्वानुमान उद्देश्यों के लिए और अधिक आसानी से इस्तेमाल किया जा सकता है क्योंकि अध्ययन चर पर प्रेक्षण की ऐतिहासिक अनुक्रम प्रकाशित माध्यमिक स्रोतों से आसानी से उपलब्ध हो जाते हैं। येक्रमानुशार प्रेक्षण सांख्यिकीय रूप से निर्भर होते हैं और काल श्रृंखला मॉडलिंग ऐसी निर्भरता के विश्लेषण के लिए तकनीकों से संबंधित है। इस प्रकार काल श्रृंखला मॉडलिंग में, भविष्य की अवधि के लिए मानों की भविष्यवाणी अध्ययन के तहत चर के पिछले मानों के पैटर्न पर आधारित है, लेकिन आम तौर पर व्याख्यात्मक चर पर नहीं जो तंत्र को प्रभावित कर सकते हैं। इस तरह के काल श्रृंखला मॉडल का सहारा लेने के दो मुख्य कारण हैं। सबसे पहला, तंत्र को समझ पाना सम्भव नहीं है, और अगर यह समझ लिया जाता है तो भी इस के कारण और प्रभाव संबंध को मापना बेहद मुश्किल हो सकता है, दूसरा, मुख्य महत्व केवल भविष्यवाणी करने के लिए हो सकता है कि क्या होगा, ना की ऐसा क्यों होता है यह जानने के लिए। कई बार, अध्ययन चर (चरों) को प्रभावित करने वाले कारण कारकों (व्याख्यात्मक चर) पर जानकारी का संग्रहण कठिन/असंभव हो सकता है और इसलिए व्याख्यात्मक चर पर लंबी श्रृंखला आंकड़ों की उपलब्धता एक समस्या है। ऐसी स्थितियों में, काल श्रृंखला के मॉडल पूर्वानुमान के लिए वरदान हैं।

काल श्रृंखला प्रक्रिया की स्थिरता

काल श्रृंखला को स्थिर (स्टेशनरी) कहा जाता है, यदि इसकी अंतर्निहित उत्पादन प्रक्रिया स्थायी माध्यम और स्थायी भिन्नता के आधार पर होती है, जो अपने ऑटोकोरिलेशन फंक्शन (ACF) के साथ अनिवार्य रूप



से समय के माध्यम से स्थिर होती है। इस प्रकार, यदि हम एक प्राप्ति के विभिन्न उपसमुच्चय (समय श्रृंखला 'नमूना') पर विचार करते हैं, तो अलग-अलग उपसमुच्चय विशेष रूप से माध्यम, वेरिएंस (variance) और ऑटोकोरिलेशन फलन में ज्यादा अन्तर नहीं होता है।

स्थिरता कि जांच के लिए डिकी फुलर टेस्ट नामक सांख्यिकीय परीक्षण का सबसे व्यापक रूप से इस्तेमाल किया जाता है। परीक्षण करने के लिए, निम्न प्रतिगमन मॉडल के अनुमान ओएलएस (OLS) के द्वारा किया जाना चाहिए—

$$y'_t = \phi y_{t-1} + b_1 y'_{t-2} + \dots + b_p y'_{t-p}$$

जहाँ y'_t डिफरेंसड श्रृंखला ($y_t - y_{t-1}$) को इंगित कर रहे हैं। प्रतिगमन में, p की संख्या आम तौर पर तीन निर्धारित कि जाती है। यदि ϕ , शून्य के पास है तो मूल श्रृंखला में डिफ्रेंसिंग की जरूरत है, अन्यथा $\phi < 0$ तो y_t पहले से ही स्थिर है।

स्वतः सहसम्बन्ध (ऑटोकोरिलेशन) फलन

(i) स्वतः सहसम्बन्ध

ऑटोकोरिलेशन का मतलब है कि एक समय श्रृंखला में प्रेक्षण एक दूसरे से संबंधित होते हैं और वर्तमान प्रेक्षण (Y_t) और वर्तमान से p अवधि पहले का प्रेक्षण (y_{t-p}) के बीच सरल सह संबंध द्वारा मापा जाता है। यह एक दी गई श्रृंखला Y_t के लिए, चलेगा (y_t) पर स्वतः सहसम्बन्ध जोड़े (y_t, y_{t-p}) के बीच सहसम्बन्ध द्वारा दिया जाता है।

$$r_p = \frac{\sum_{t=1}^{n-p} (Y_t - \bar{Y})(Y_{t+p} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2}$$

इसकी सीमा -1 से 1 होती है। बॉक्स और जेन किन्स ने सुझाव दिया कि अधिकतम उपयोगी r_p की संख्या लगभग $N/4$ होनी चाहिए, जहाँ N उस अवधि की संख्या है, जिस पर Y_t की जानकारी उपलब्ध है।

(ii) आंशिक स्वः सहसम्बन्ध

आंशिक स्वः सहसंबन्ध y_t और y_{t-p} के मध्य सहयोग की डिग्री को मापने के लिए उपयोग किया जाता है जब समय लेग पर y -प्रभाव $1, 2, 3, \dots, p-1$ को हटा दिया जाता है।

(iii) स्वः सहसम्बन्ध फलन (एसीएफ) तथा आंशिक स्वः सहसम्बन्ध फलन (पीएसीएफ)

सैद्धांतिक रूप से एसीएफ और पीएसीएफ (ऑटोकोरिलेशन बनाम लेग) विभिन्न मॉडलों ऑटोरेग्रेसिव और चल औसत तत्वों (मुविंग एवरेज कॉम्पोनेंट) के विभिन्न ऑर्डर के मान के लिए, अर्थात् p और q के लिए उपलब्ध हैं (देखें, पंक्रट्स, 1984). इस प्रकार इन सैद्धांतिक ए सी एफ / पी ए सी एफ के साथ दिए गए समय श्रृंखला आंकड़ों से प्राप्त कोरलोग्राम (ए सी एफ बनाम लेग) की तुलना करें, एक अच्छा मिलान खोजने के लिए और वैकल्पिक रूप से एक या अधिक अरीमा (ARIMA) मॉडल का चयन करें। सैद्धांतिक एसीएफ और पीएसीएफ के सामान्य लक्षण निम्नानुसार हैं (यहां 'स्पाइक' स्वतः सहसम्बन्ध के परिमाण के बराबर लम्बाई (length) वाली विभिन्न लेग की रेखा को दर्शाता है)

मॉडल	ए सी एफ	पी ए सी एफ
ए आर (AR)	स्पाइक क्षय शून्य की ओर	स्पाइक को शून्य से छांटना (cutoff)
एम ए (MA)	प्रवाह को शून्य छांटना (cutoff)	स्पाइक क्षय शून्य की ओर
आरमा (ARMA)	प्रवाह की क्षय शून्य की ओर	स्पाइक क्षय शून्य की ओर

एरिमा निरूपण का विवरण

(i) एरिमा (ARIMA) मॉडलिंग

सामान्य तौर पर, एक अरिमा मॉडल को अरिमा (p,d,q) द्वारा चिन्हित किया जाता है, जहाँ p, d तथा q क्रमशः स्वतः प्रतिगमन, एकीकरण (डिफरेंसिंग) तथा चल औसत के ऑर्डर को निरूपित करते हैं। अरिमा भाषा में, समय श्रृंखला पिछले वास्तविक मानों और



यादृच्छिक झटके का एक रेखिक फलन है। उदाहरण के लिए, एक समय श्रृंखला प्रक्रिया $\{y_t\}$ दी गई है, पहली ऑर्डर स्वतः प्रतिगमन प्रक्रिया को ARIMA (1,0,0) या AR (1) द्वारा चिन्हित किया गया हैजिसे निम्न द्वारा दिया जाता है।

$$y_t = \mu + \phi_1 y_{t-1} + \varepsilon_t$$

और पहला ऑर्डर चल औसत प्रक्रिया को ARIMA (0,0,1) या साधारणतया MA (1) द्वारा निर्दिष्ट है और इसे निम्न द्वारा दिया जाता है।

$$y_t = \mu - \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

वैकल्पिक रूप से, अंततः व्युत्पन्न मॉडल, ये प्रक्रियाओं और उच्चतर ऑर्डर का मिश्रण भी हो सकता है।

इस प्रकार एक स्थिर IRMA (p, q) प्रक्रिया समीकरण द्वारा परिभाषित की गई है।

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} + \dots - \theta_q \varepsilon_{t-q} + \varepsilon_t$$

जहाँ ε_t स्वतंत्र रूप से और जो सामान्य रूप से शून्य माध्य और समान विभिनन्ता σ^2 के लिए $t = 1, 2, \dots, n$ के साथ normal distribution को follow करता है। यहाँ ध्यान दें कि प्रयोग में p और q कामान शून्य और तीन के बीच होता है। (मुख्य चर के डिफरेंसिंग की डिग्री सेक्शन 5 (i) में चर्चा की जाएगी)

(ii) ऋतु-संबंधी एरिमा मॉडलिंग

प्रासंगिक मॉडलों की पहचान और उपयुक्त ऋतु-संबंधी परिवर्तनों को शामिल करना ऋतु-संबंधी मॉडल के लिए आवश्यक है।

ऋतु-संबंधी एरिमा जैसेकि ARIMA (p, d, q) (P,D,Q) मॉडल को निम्न तरीके से परिभाषित किया जाता है

$$\phi_p(B)\theta_q(B^s)\nabla^d \nabla_s^D y_t = \theta_q(B^s)\theta_s(B)\varepsilon_t$$

जहाँ

$$\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p, \theta_s(B) = 1 - \theta_1 B - \dots - \theta_s B^s$$

$$\theta_p(B^s) = 1 - \theta_1 B^s - \dots - \theta_p B^{sp}, \Theta_q(B^s) = 1 - \Theta_1 B^s - \dots - \Theta_q B^{sq}$$

B बैक शिफ्ट ऑपरेटर है, (जैसे की $B y_t = y_{t-1}$, $B^2 y_t = y_{t-2}$ और इसी तरह) 's' ऋतु-संबंधी लेग और 'et' माध्य 0 और विचरण σ^2 के साथ स्वतंत्र normal सत्रु टिचर का अनुक्रम है। ϕ 's और θ 's के क्रमशः ऋतु-संबंधी और गैर ऋतु-संबंधी स्वतः प्रतिगमन मापदंड हैं। Θ 's और θ 's के क्रमशः ऋतु-संबंधी और गैर-ऋतु-संबंधी चल औसत मापदंड हैं। p और q क्रमशः गैर ऋतु-संबंधी स्वतः प्रतिगमन और चल औसत मापदंड के ऑर्डर है, जब कि P और Q क्रमशः ऋतु-संबंधी स्वतः प्रतिगमन तथा चल औसत मापदंड हैं। d और D क्रमशः औसत गैर ऋतु-संबंधी और ऋतु-संबंधी अंतर को दर्शाते हैं।

एरिमा मॉडल निर्माण की कला

(i) पहचान

मॉडलिंग की प्रक्रिया में सबसे महत्वपूर्ण कदम सीरीज की स्थिरता की जांच करना है, क्योंकि आकलन प्रक्रियाएं केवल स्थिर श्रृंखला के लिए उपलब्ध हैं। यहाँ दो प्रकार की स्थिरता होती है अर्थात माध्य में स्थिरता और भिन्नता में स्थिरता। आंकड़ों के ग्राफ और स्वतः सहसम्बंध की संरचना तथा आंशिक सह संबंध गुणांक पर एक सरसरी नजर डालें तो स्थिरता की उपस्थिति के लिए सुराग प्रदान कर सकते हैं। स्थिरता की जांच करने का एक अन्य तरीका यह है कि मूल आंकड़ों के लिए एक प्रथम ऑर्डर स्वतः प्रतिगमन मॉडल फिट करना है और यह जांच करना है कि ϕ_1 गुणांक एक से कम है या नहीं। यदि मॉडल अस्थायी पाया जाता है तो अधिकतर स्थिरता श्रृंखला को विभाजित करके प्राप्त की जा सकती है या डि की फुलर टेस्ट के लिए जाएं (अनुभाग 2 देखें)।

यह गैर ऋतु-संबंधी और ऋतु-संबंधी दोनों के लिए लागू होता है। इस प्रकार, यदि 'x t' मूल श्रृंखला को दर्शाता है, तो पहले क्रम का गैर ऋतु-संबंधी का अंतर है—

$$Y_t = X_t - X_{t-1}$$

ऋतु-संबंधी अंतर के बाद (यदि आवश्यक हो)

$$Z_t = Y_t - Y_{t-s} = (X_t - X_{t-1}) - (X_{t-s} - X_{t-s-1})$$



पहचान प्रक्रिया में अगला कदम गैर ऋतु-संबंधी और ऋतु-संबंधी मापदंडों, p, q और P, Q के क्रमों के लिए प्रारंभिक मानों को ढूँढना है। वे सार्थक स्वतः सहसम्बंध और आंशिक स्वतः सहसम्बंध के गुणांक देखकर प्राप्त किए जा सकते हैं (अनुभाग 3 (iii) देखें)। जैसे कि, यदि दूसरा क्रम स्वतः सहसम्बंध गुणांक सार्थक है, तब एक AR (2), या MA (2) या ARMA (2) मॉडल के साथ शुरू करने की कोशिश की जा सकती है। यह एक कठोर नियम नहीं है, क्योंकि नमूना स्वतः सहसम्बंध गुणांक मूल स्वतः सहसम्बंध गुणांक का कमजोर अनुमान हैं। फिर भी इन्हें प्रारंभिक मान के रूप में इस्तेमाल किया जा सकता है, जबकि बार-बार चरणों के बाद अंतिम मॉडलों को हासिल किया जाता है। ध्यान दें कि आमतौर पर p, d , या q के लिए दो क्रम तक अभ्यास में एक अच्छा मॉडल विकसित करने के लिए पर्याप्त हैं।

(ii) आकलन

पहचान चरण में एक या एक से अधिक मॉडलों को वैकल्पिक रूप से चुना जाता है जो उपलब्ध डेटा को सांख्यिकीय रूप से पर्याप्त प्रतिनिधित्व करते हैं। तब हम कम से कम वर्गों (लीस्ट स्क्वेयर) के मॉडल के मापदंडों के सटीक अनुमान प्राप्त करने का प्रयास करते हैं, जैसा कि बॉक्स और जेनकिंस द्वारा सुझाव दिया गया। मानक कंप्यूटर पैकेज जैसे कि एस ए एस, एस पी एस एस, आर इत्यादि उपयुक्त पद्धतियों का उपयोग कर प्रासंगिक मानकों के अनुमानों को प्राप्त करने के लिए उपलब्ध हैं। संक्षेप के लिए आकलन के तरीकों पर यहां चर्चा नहीं की गई है।

(iii) निदान

AR और MA के विभिन्न संयोजनों के लिए व्यक्तिगत और सामूहिक रूप से विभिन्न मॉडलों को प्राप्त किया जा सकता है। निम्नलिखित निदान के साथ सबसे अच्छा मॉडल प्राप्त किया जाता है—

(अ) अकाईक सूचना मापदंड (AIC)/बेइसियन सूचना मापदंड (BIC)/श्वार्ज-बेइसियन सूचना मापदंड (SBC)

एआइसी, $AIC = (-2 \log L + 2 m)$ द्वारा दिया जाता है, जहां $m = p + q + P + Q$ और L लाइकलीहूड फलन है। चूंकि $-2 \log L$ लगभग $(n = 1 + \log 2\pi) + n \log \sigma^2$ के बराबर है, जहां σ^2 मॉडल MSE है, AIC को लिखा जा सकता है $AIC = \{n(1 + \log 2\pi) + n \log \sigma^2 + 2 m\}$ और क्योंकि इस समीकरण में पहली पद स्थिर है, यह आमतौर पर मॉडल के बीच तुलना करते समय छोड़ा जाता है। AIC के लिए एक विकल्प के रूप में कभी-कभी SBC भी उपयोग किया जाता है जो निम्न द्वारा दिया जाता है $SBC = \log \sigma^2 + (m \log n) / n$.

(ब) अवशिष्ट ए सी एफ का प्लॉट

एक बार उचित एरिमा मॉडल फिट होने के बाद, फिटेड मॉडल के अवशिष्ट ए सी एफ के प्लॉट करके गुडनेस ऑफ फिट की जांच कर सकता है। यदि अवशेषों के अधिकांश नमूनों का स्व-संबंध गुणांक $\pm 1.96 / \sqrt{N}$ सीमाओं के भीतर हैं, जहां N उन प्रेक्षणों की संख्या है, जिस पर मॉडल आधारित है, तब अवशिष्ट व्हाइट नोइस हैं जो दर्शाता है कि मॉडल एक अच्छा फिट है।

(स) पोर्टमोंट्यू परीक्षणों के माध्यम से अवशेषों के ऑटो सहसंबंधों का असार्थक (काई स्क्वेयर टेस्ट पर आधारित Q-टेस्ट) बॉक्स-पियर्स या जंग-बॉक्स टेस्ट

आंकड़ों को अस्थायी मॉडल में फिट करने के बाद, मॉडल की पर्याप्तता का परीक्षण करने के लिए नैदानिक जांच करना महत्वपूर्ण है और यदि आवश्यक हो तो संभावित सुधारों का सुझाव देने के लिए। यह पूरा करने का एक तरीका अवशिष्टों के विश्लेषण के माध्यम से है। यह पाया गया है कि बॉक्स-पियर्स स्टैटिस्टिक्स (अवशेषों के स्वतः संबंधों का एक फलन) के रूप में जाना जाने वाला मात्रा Q का परीक्षण करके चुने हुए मॉडल की समग्र पर्याप्तता को मापने के लिए प्रभावी होता है। जिसका अनुमानित वितरण



काई-स्क्वेयर है और जिसकी निम्नानुसार गणना की जाती है:

$$Q = n \sum r^2(j)$$

जहाँ योग 1 से k तक है तथा k अधिकतम लेग माना गया है, श्रृंखला में प्रेक्षकों की संख्या n है, $r(j)$, lag में अनुमानित स्वतः सम्बन्ध है k कोई भी सकारात्मक पूर्णांक हो सकता है जो आमतौर पर लगभग 20 होता है $AQ(k-m_1)$ स्वतंत्रता की डिग्री के साथ काई - स्क्वेयर टेस्ट का अनुगमन करता है जहाँ मॉडल में अनुमानित मापदंडों की संख्या m_1 है। एक संशोधित Q स्टैटिस्टिक जंग-बॉक्स स्टैटिस्टिक है जो कि निम्न द्वारा दिया जाता है:

$$Q = n(n+2) \sum r^2(j) / (n-j)$$

Q स्टैटिस्टिक की काई-स्क्वेयर वितरण की महत्वपूर्ण मानों (critical values) से तुलना की जाती है। यदि मॉडल सही ढंग से निर्दिष्ट है, तो अवशिष्ट असहसम्बन्ध होना चाहिए और Q छोटा होना चाहिए

(संभाव्यता मान बड़ा होना चाहिए)। एक सार्थक मान इंगित करता है कि चुना हुआ मॉडल अच्छी तरह से उपयुक्त नहीं है।

इन सभी चरणों को काफी देखभाल और कार्य की आवश्यकता होती है और वे स्वयं परिपूर्ण नहीं होते हैं।

संदर्भ

1. बॉक्स, जी.ई.पी., जेनकिन्स, जी.म. एवं रैनसेल, जी.सी. (1994). समय श्रृंखला विश्लेषण: पूर्वानुमान और नियंत्रण, पियरसन शिक्षा, दिल्ली.
2. मेक्रीडाकिस, स., व्हीलवराइट, स.सी. एवं हिंडमान, आर.जे. (1998)। पूर्वानुमान विधियाँ और अनुप्रयोग, 3 वी संपादन, जॉन विले, न्यूयॉर्क.
3. पेनक्रात्ज, ए. (1983). फोरकास्टिंग विद यूनिवर्सिटी बॉक्स - जेनकिन्स मॉडल्स: कन्सेप्ट्स और केस, जॉन विले, न्यूयॉर्क.

ए.एन.एन. का अवलोकन: उत्तर प्रदेश के झांसी जिले की मानसून वर्षा की भविष्यवाणी में अनुप्रयोग

मुन्मय राय, के. एन. सिंह, प्रकाश कुमार, राजीव रंजन कुमार, संतोष राठौड़,
प्रदीप बसाक, हिमाद्रि शेखर राय एवं अचल लामा

आर्टिफिशियल न्यूरल नेटवर्क (ANN) मानव मस्तिष्क का कंप्यूटेशनल मॉडल पर आधारित है, जिसमें बहुत सारे साधारण प्रोसेसिंग यूनिट्स (कृत्रिम न्यूरॉन्स) एक दूसरे से जुड़े होते हैं एवं काफी संख्या में वेटेड कॉनेक्शंस के तहत एक दूसरे से सूचना का आदान प्रदान करते हैं। मानव मस्तिष्क नियूरान्स से बना होता है। साइनप्सेस के तहत एक न्यूरॉन दूसरे न्यूरॉन से इनपुट प्राप्त करता है। आर्टिफिशियल न्यूरल नेटवर्क का उपयोग पैटर्न क्लासिफिकेशन, लर्निंग के तहत पैटर्न रिकग्निशन इत्यादि में किया जाता है। ANNs नॉन लीनीयर, डाटा ड्रिवन एवं सेल्फ अडैप्टिव अप्रोच है। इस तकनीक की सहायता से बिना इनपुट एवं आउटपुट के बीच संबंध जाने, नॉन लीनीयर मॉडलिंग बहुत आसानी के किया जा सकता है। इस तकनीक का उपयोग मेडिसिन, इंजीनियरिंग, एकानमिक्स, जियोलॉजी, फिजिक्स, कृषि इत्यादि में किया जाता है।

न्यूरल नेटवर्क

न्यूरल नेटवर्क एक दूसरे से जुड़े हुए कनेक्टेड सेल्स (न्यूरॉन) का एक समूह है। न्यूरल नेटवर्कस न्यूरॉन्स के लेयर्स से बने होते हैं जो इस तरह से जुड़े होते हैं की एक लेयर पहले वाले न्यूरॉन्स के लेयर से इनपुट प्राप्त करता है एवं आउटपुट को अगले वाले लेयर को भेज देता है। न्यूरन इनपुट वेक्टर (y_1, y_2, \dots, y_k) का वास्तविक फंक्शन होता है, जिसका आउटपुट निम्न तरीके से प्राप्त किया जाता है:

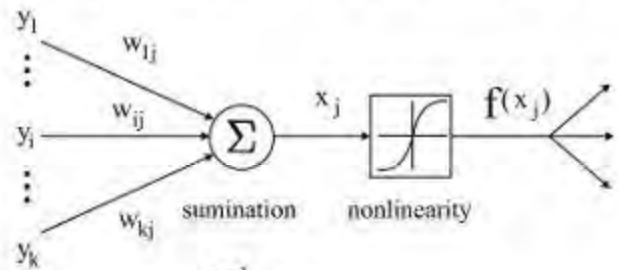
$$f(x_j) = f(a_j + \sum_{i=1}^k w_{ij} y_i)$$

जहाँ f एक फंक्शन है जैसे की sigmoid (logistic or tangent hyperbolic) फंक्शन।

न्यूरल नेटवर्क का स्किमेटिक एवं मैममेटिकल प्रारूप आकृति 1 एवं आकृति 2 क्रमशः में दिए गये हैं।



आकृति 1: न्यूरल नेटवर्क का स्किमेटिक प्रारूप



$$x_j = \sum_{i=1}^k w_{ij} y_i$$

आकृति 2: साधारण न्यूरल नेटवर्क

न्यूरल नेटवर्कस का आर्किटेक्चर

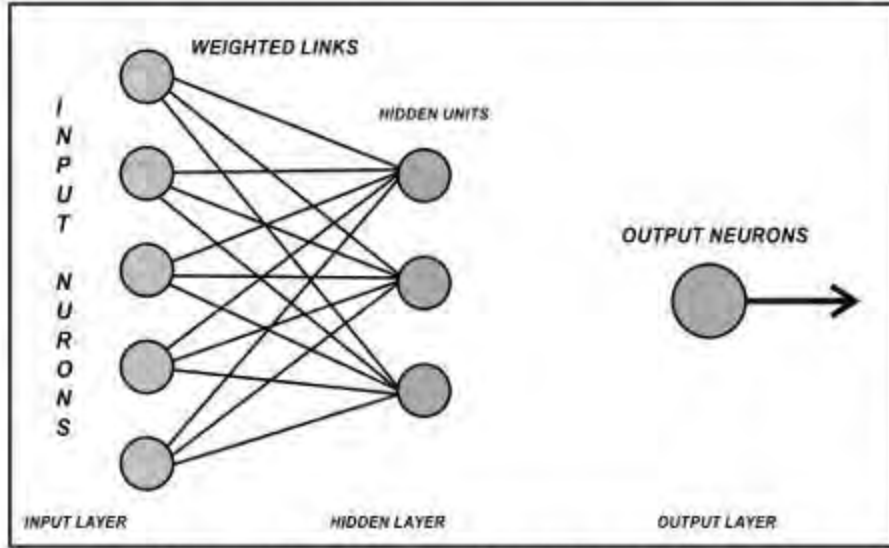
सबसे अधिक उपयोग में आने वाले आर्किटेक्चर के बारे में नीचे दिए गये हैं:

फीड फॉरवर्ड न्यूरल नेटवर्कस

इस नेटवर्क में, एक दूसरे से जुड़े हुए यूरान के तहत सूचना इनपुट लेयर से हिडन लेयर होते हुए अंत में आउटपुट लेयर की ओर जाता है।

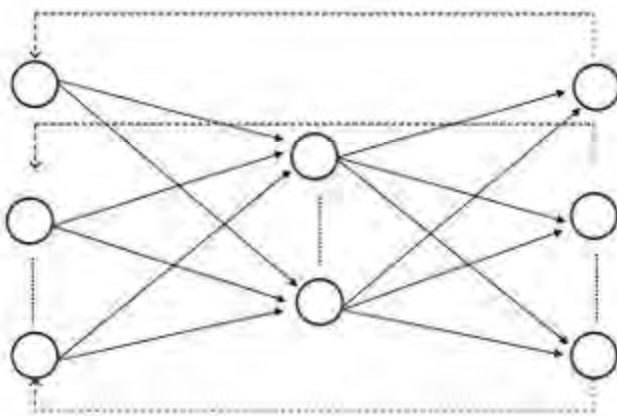
रिकरेंट न्यूरल नेटवर्कस

यह नेटवर्क, फीड फॉरवर्ड नेटवर्क से भिन्न है, इस नेटवर्क में कम से कम एक फीड बैक लूप होता है। आकृति 3 में रिकरेंट न्यूरल नेटवर्कस का ग्राफिकल प्रारूप दिया गया है।



SCHEMATIC REPRESENTATION OF A MULTILAYER FEED FORWARD NEURAL NETWORK (MLP)

आकृति 3: मल्टी लेयर फीड फॉरवर्ड न्यूरल नेटवर्क

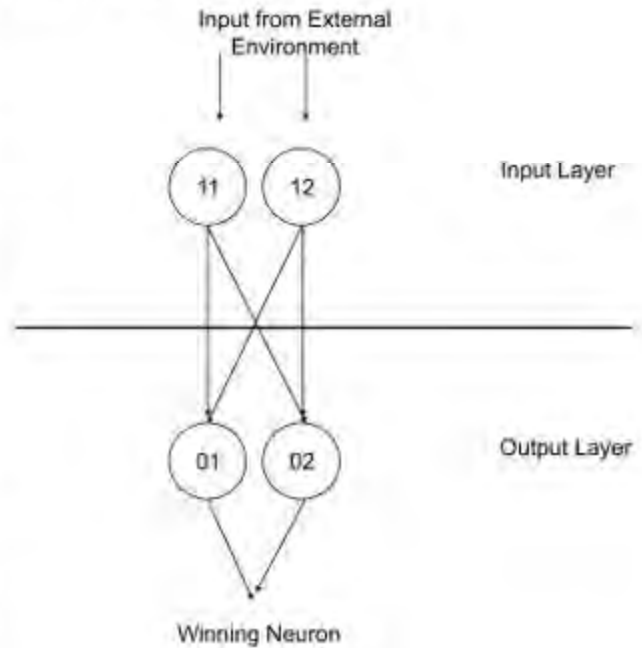


आकृति 4: रिकरेंट न्यूरल नेटवर्कस का ग्राफिकल प्रारूप

RBF नेटवर्कस भी फीड फॉरवर्ड की तरह होता है लेकिन RBF में केवल एक हिडन लेयर होता है। जिसमें इनपुट्स की संख्या कुछ भी हो सकता है और अक्सर हिडन लेयर में रेडियल काम्बिनेशन फंक्शंस उपयोग किए जाते हैं, जो की इनपुट वेक्टर और वेट वेक्टर के बीच स्क्वेर्ड यूक्लीडेआन डिस्टेन्स के आधार पर निर्धारित किया जाता है।

कोहोनेन न्यूरल नेटवर्कस

कोहोनेन या SOFM नेटवर्कस का उपयोग दूसरे नेटवर्कस के साथ बिल्कुल भिन्न तरीके से किया जाता है। दूसरे नेटवर्कस सुपरवाइज्ड लर्निंग के



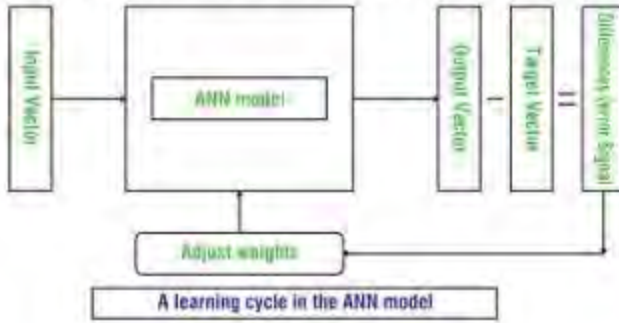
आकृति 5: कोहोनेन न्यूरल नेटवर्क

लिए डिजाइन किए गये हैं जबकि SOFM नेटवर्कस अन-सुपरवाइज्ड लर्निंग के लिए डिजाइन किया गया है।

SOFM नेटवर्कस में केवल दो लेयर्स होते हैं: रेडीयल यूनिट्स का इनपुट और आउटपुट लेयर्स (जो की टॉपोलॉजिकल मैप लेयर के नाम से जाना जाता है)।

लर्निंग तकनीकें

न्यूरल नेटवर्कस में लर्निंग तकनीकों को तीन प्रकार से परिभाषित किया जा सकता है: सुपरवाइज्ड, अन-सुपरवाइज्ड एवं री-इन्फोर्सड।



आकृति 6 :

सुपरवाइज्ड लर्निंग

इस तकनीक में, प्रत्येक इनपुट पैटर्न जो की नेटवर्क को ट्रेन करने में उपयोग किया जाता है, आउटपुट पैटर्न से जुड़ा होता है, जो की हमारा लक्ष्य है। एरर कंप्यूट करने के लिए, नेटवर्कस आउटपुट और करेक्ट एक्सपेक्टेड आउटपुट की तुलना करते हैं। कंप्यूटेड एरर्स का उपयोग नेटवर्कस के पैरामीटर्स बदलने में करते हैं, जोकि तकनीक का परफॉर्मन्स को बढ़ाता है।

अन-सुपरवाइज्ड लर्निंग

इस तकनीक में, लक्ष्य वाले आउटपुट को नेटवर्क में प्रेजेंट नहीं करते हैं। इस तकनीक में सिस्टम्स खुद ब खुद इनपुट पैटर्नस में स्ट्रक्चरल पैटर्नस की खोज कर के सीखता है।

री-इन्फोर्सड लर्निंग

इस तकनीक में टीचर तो होता है लेकिन संभावित उत्तर के बारे में कुछ भी नहीं बताता है, केवल यह बताता है की कंप्यूटेड आउटपुट सही है या गलत है। उपलब्ध कराए गये सूचना लीनिंग प्रोसेस में नेटवर्कस की मदद करता है, जिसमें सही उत्तर कंप्यूट करने पर इनाम दिया जाता है एवं गलत उत्तर पर पेनाल्टी दी जाती है।

ANN मॉडल बनाना

मॉडलिंग के बहुत सारे इसूज हैं जो ANN की परफॉर्मन्स को प्रभावित करते हैं, उनमें से एक है उपयुक्त आर्किटेक्चर को निर्धारित करना जैसे की लेयर्स की संख्या, प्रत्येक लेयर में नोड्स की संख्या एवं arcs की संख्या जो नोड्स को एक दूसरे से जोड़ते हैं। अन्य नेटवर्क डिजाइन का निर्णय जैसे की हिडन और आउटपुट नोड्स के आक्टिवेशन फन्क्शन्स को चुनना, ट्रेनिंग अल्गोरिथ्म, डाटा ट्रॉन्सफॉर्मेशन, ट्रेनिंग एवं टेस्ट्स सेट्स और परफॉर्मन्स मिजर।

ANN मॉडल बनाने स्टेप्स

- वेरिबल्स का चुनाव करना
- ट्रेनिंग, टेस्टिंग एवं वेलिडेशन सेट्स का बनाना
- नेटवर्क स्ट्रक्चर
- हिडन लेयर्स एवं नोड्स की संख्या
- आउटपुट नोड्स की संख्या
- नोड्स का आपसी जुड़ाव

आक्टिवेशन फंक्शंस

आक्टिवेशन फंक्शंस को ट्रान्सफर फंक्शंस के नाम से जाना जाता है। यह इनपुट एवं आउटपुट के नोड्स एवं नेटवर्क के बीच निकटता निर्धारित करता है। अक्सर, आक्टिवेशन फंक्शंस नॉन-लीनियरिटी का डिग्री दर्शाता है जो की ANN के लिए अत्यंत महत्वपूर्ण है और वो निम्नलिखित हैं:

- i. सिगमोएड (लजिस्टिक) फंक्शन
- ii. हाइपरबॉलिक टैनजेंट (tanh) फंक्शन
- iii. Sine or Cosine फंक्शन
- iv. लीनियर फंक्शन

हिडन एवं आउटपुट नोड्स के लिए उपर दिए गये आक्टिवेशन फंक्शंस में से लॉजिस्टिक ट्रान्सफर फंक्शन सबसे अधिक उपयोग किया जाता है। जबकि, फोरकोस्टिंग से जुड़े समस्या के लिए आउटपुट नोड्स के लिए लीनियर ट्रान्सफर फंक्शन का उपयोग किया जाता है।



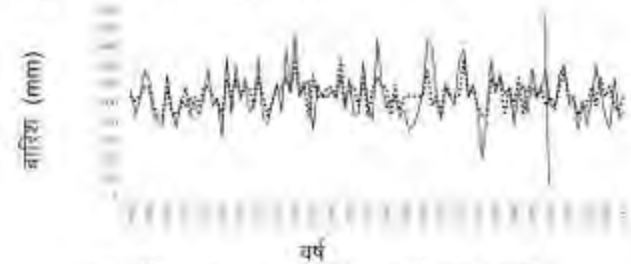
एवॉल्यूशन क्राइटेरिया

नियुरल नेटवर्कस में सम ऑफ स्क्वर्ड एरर सबसे अधिक मिनिमाइज किया जाने वाला एरर फंक्शन है। लिस्ट अएब्सोल्यूट डीवियेशन, लिस्ट फोर्थ पॉवर, असिमेट्रिक लिस्ट स्क्वेर एवं पर्सेंटेज डिफरेंसेस इत्यादि अन्य एरर फंक्शंस हैं।

उदाहरण

1901 से 2013 के बीच उत्तर प्रदेश के झांसी जिले के मासिक वर्षा (मिमी) के आंकड़े, भारतीय उष्णकटिबंधीय मौसम विज्ञान, पुणे, भारत की साइट (www-tropmet-res-in) पर उपलब्ध है, लिया गया है। इस आंकड़ों से, हर महीने के लिए मानसून वर्षा डेटा चार महीनों के दौरान मासिक वर्षा डेटा जोड़कर हासिल किया गया है, अर्थात् जून से सितंबर तक। मॉडल विकास के लिए 1901 से 2003 तक का उपयोग किया गया है और 2004 से 2013 का पूर्वानुमान उद्देश्य के लिए उपयोग किया गया है।

1901 से 2003 की अवधि के दौरान झांसी के मानसून की बारिश की श्रृंखला का प्लॉट आकृति 7 में दिया गया है।



आकृति 7: मूल समय श्रृंखला डेटा का प्लॉट

काल श्रृंखला की प्लाट का पता चलता है कि समय के साथ एक अव्यवस्थित पैटर्न है जो समय श्रृंखला को प्रकृति में अरैखीय दर्शाता है।

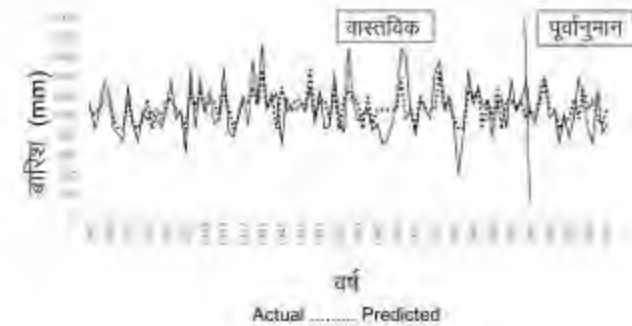
पहले मूल समय श्रृंखला की जाँच करते हैं तदोपरान्त डेटा में ANN लगाते हैं। चूंकि यहाँ p, q, का मान लेने के लिए कोई सिद्ध सिद्धांत नहीं है इसलिए, हमने अंतराल और छिपी हुई यूनिट के विशिष्ट संयोजन का प्रयास किया है। 10 लर्निंग की दर और गति के विशिष्ट संयोजन का प्रयास

किया गया है। एएनएन 5 लेग और 6 छिपी परत के साथ दूसरे संयोजन से बेहतर प्रदर्शन देखा गया। 0.02 मुमेंटम के साथ लर्निंग की दर 0.04 सर्वश्रेष्ठ परिणाम देता है। पुनरावृत्तियों की कुल संख्या 1000 लिया गया है। ANN मॉडल की रूपरेखा तालिका -1 में नीचे दी गई है।

तालिका 1: वर्षा डेटा के लिए ANN पैरामीटर

पैरामीटर	ANN
Number of input (lag)	5
Number of hidden unit	6
Activation function in hidden unit	Sigmoid
Number of iterations	1000
Learning algorithm	Gradient decent back propagation
Learning rate	0.04
Momentum	0.02

फिट और पूर्वानुमानित मानों के चित्रमय चित्रण को नीचे आकृति 8 में दिया गया है।



आकृति 8: वास्तविक बनाम मॉडल मान

आकृति 8 से यह समझा जा सकता है कि मॉडल अच्छा फिट है।

निष्कर्ष

ANN अक्सर बारिश की भविष्यवाणी, कृषि मूल्य आदि की भविष्यवाणी के क्षेत्र में गैर-अक्षीय मॉडलिंग के रूप में इस्तेमाल किया जा सकता है।



संदर्भ

1. गीता, जी एवं सेल्वराज, आर. एस. 2011. प्रिडिक्शन ऑफ मंथली रैनफॉल डाटा इन चेन्नई उजिंग ब्लाक प्रोपोगेशन न्यूरल नेटवर्क मॉडल इन्टरनेशनल जर्नल ऑफ इंजीनियरिंग साइंस ऐंड टेक्नोलोजी, 3: 211-213.
2. गुहथाकुर्ता, पी. 2006. लॉन्ग-रेंज मानसून रेनफॉल प्रिडिक्शन ऑफ 2005 फॉर दि डिस्ट्रिक्ट ऐंड सब-डिविजन ऑफ केरला विद आर्टिफिशियल न्यूरल नेटवर्क. करेंट साइंस, 90: 773-779.
3. हंग, एन. क्यू., बाबेल, एम. एस., विस्कल, एस. एवं त्रिपाठी एन. के. 2008. एन आर्टिफिशियल न्यूरल नेटवर्क मॉडल फॉर रेनफॉल फोरकास्टिंग इन बैंकाक, थाईलैंड. हाई ड्रोलोजी ऐंड अर्थ साइंसेज, 5: 183-218.
4. लीन, जी. एवं चेन, एल. 2005. एप्लिकेशन ऑफ एन आर्टिफिशियल न्यूरल नेटवर्क टू टाईफून रेनफॉल फोरकास्टिंग ड्रोलोजिकल प्रोसेस, 19: 1825-1837.

एस.ए.एस. द्वारा आंकड़ों की सारांश रिपोर्ट

राजेंद्र सिंह तोमर एवं प्रमोद कुमार

परिचय

आंकड़ों की प्रस्तुति एवं विश्लेषण प्रत्येक अनुसंधान अध्ययन का एक अनिवार्य हिस्सा है। एस ए एस सॉफ्टवेयर के अन्दर कई साधन व विधियाँ हैं जो कि सॉफ्टवेयर उपयोगकर्ता को विभिन्न प्रकार के चार्ट, रिपोर्ट और आंकड़ों का सारांश रिपोर्ट तैयार करने में उदारता प्रदान करते हैं। सभी साधनों में, PROC PRINT PROC REPORT व PROC TABULATE विशेष रूप से शक्तिशाली और बहुमूल्य साधन हैं जिसका प्रयोग आंकड़ों की सारांश रिपोर्ट तैयार करने में किया जा सकता है। इनमें से PROC REPORT एस ए एस का बहुत अहम् हिस्सा है। बायोलॉजिकल अन्वेषक, अपने अन्वेषण अध्ययन को सांख्यिकी विवेचना के माध्यम से सरल एवं अधिक उपयोगी बनाने के लिए PROC REPORT का प्रयोग बहुत आसानी से कर सकते हैं। इस लेख अथवा दस्तावेज के माध्यम से आंकड़ों की विभिन्न प्रकार की रिपोर्ट तैयार करने के लिए PROC PRINT PROC REPORT व PROC TABULATE अति उपयोगी कथनों का प्रयोग कर, प्रयोगकर्ता के लिए सुलभ साधन से परिचय करने का प्रयास किया गया है।

PROC PRINT द्वारा साधारण रिपोर्ट (Simple Report Using PROC PRINT)

साधारण रिपोर्ट तैयार करने के लिए PROC PRINT का प्रयोग सबसे आसान है। सामान्यतः एस ए एस प्रोग्राम में PROC REPORT कथन का प्रयोग आंकड़ों को देखने या जाँचने के लिए किया जाता है। इस कथन का उपयोग आंकड़ों को जाँचने के बावजूद छोटों छोटी गणना के साथ एक रिपोर्ट तैयार करने के लिए भी कर सकते हैं जो इसकी उपयोगिता को प्रदर्शित करेगा। इस उपयोगिता को स्पष्ट करने के लिए हम एक छोटा सा उदाहरण प्रस्तुत करते हैं जिसमें चार वर्षों

की पैदावार (Yield), उत्पादन लागत (Production Cost) और प्रतिफल (Return) के आंकड़े लिए गए हैं। सबसे पहले एस ए एस के साधारण कथनों के माध्यम से एक डाटा सेट तैयार करते हैं को इस प्रकार है

```
data rep01;
input year yld ret pc;
cards;
1991 892 13380 10021
1991 912 13680 11928
1993 1098 16020 14600
1994 1043 15645 15100
```

यहाँ पर चरों के नाम year वर्ष, yld पैदावार, ret प्रतिफल व pc उत्पादन लागत के लिए प्रयोग किए गए हैं। यदि हम इन कथनों के प्रस्तुत (submit) कर देंगे तो एसएसएस का डाटा सेट rep01 के नाम से तैयार हो जाएगा लेकिन उपयोगकर्ता सीधे इस डाटा सेट को नहीं देख पाएगा कि आंकड़ों के त्रुटि की जांच कर पायेंगे। अब हम ऊपर लिखे कोड के नीचे, निम्न पंक्तियाँ और जोड़ देते हैं जिसे प्रस्तुत करते ही, हमें शीर्षक के साथ सम्पूर्ण डाटा सेट, जांच के लिए स्क्रीन पर दिखाई देगा।

```
proc print data=rep01;
title 'Simplest Report Using PROC PRINT';
run;
```

एसएसएस प्रोग्राम में बड़ा डाटा सेट होने की दशा में, यह जांच की एक सामान्य प्रक्रिया है। अब मान लेते हैं कि हमें एक नए चर (variable), लाभ (Profit) की गणना कर, आंकड़ों की साथ रिपोर्ट प्रस्तुत करनी है इसके लिए हम उपरोक्त प्रोग्राम में, लाभ गणना के



लिए, INPUT कथन के नीचे एक कथन profit=ret-pc जोड़ते हैं। अब हमारा प्रोग्राम इस प्रकार होगा :-

```
data repo1;
input year yld ret pc;
profit=ret-pc;
cards;
1991 892 13380 10021
1991 912 13680 11928
1993 1098 16020 14600
1994 1043 15645 15100
;
run;
proc print data=repo1;
title 'Simplest Report Using PROC PRINT';
run;
```

इस प्रोग्राम को प्रस्तुत करने पर हमें निम्न रिपोर्ट प्राप्त होगी।

Simplest Report Using PROC PRINT					
Obs	year	yld	ret	pc	profit
1	1991	892	13380	10021	3359
2	1991	912	13680	11928	1752
3	1993	1098	16020	14600	1420
4	1994	1043	15645	15100	545

रिपोर्ट को और अधिक सूचनाप्रद बनाने के लिए हम PROC REPORT का एक नया प्रोग्राम तैयार करते हैं जिसमें चरों के नाम के स्थान पर मद का पूरा नाम आए तथा वर्षवार सभी मदों के आंकड़ों के साथ उनका योग भी प्रदर्शित हो।

```
proc print data=repo1 split='*';
label yld=' YIELD (in Q)'
ret='RETURN (in Rs)'
pc='PRODUCTION COST (in Rs)
profit='Profit Before *Taxes';
id year;
```

sum yld ret pc profit;

title 'PROC PRINT with Totals, Formats, Labels, and ID Variable;

split= "*" परिणाम रिपोर्ट में, स्तम्भ के शीर्षकों के नाम को दो या अधिक स्तम्भ पंक्तियों में आवश्यकता के अनुसार स्वतः विभक्त कर देगा। label चरों के नाम को परिणाम ने उसी प्रकार लिखेगा जैसे कि हमने परिभाषित किया है। id year वर्षवार सभी चरों के वास्तविक मान को प्रदर्शित करेगा और sum चरों के मान कि गणना कर योग निकाल कर प्रदर्शित करेगा। प्रस्तुत करने पर निम्न परिणाम प्राप्त होता है

PROC PRINT with Totals, Formats, Labels, and ID

PRODUCTION				
Year	YIELD (in Q)	RETURN (in Rs)	COST (in Rs)	Profit Before Taxes
1991	892	13380	10021	3359
1992	912	13680	11928	1752
1993	1098	16020	14600	1420
1994	1043	15645	15100	545
	3945	58725	51649	7076

हम दो या अधिक चरों की तुलना भी आसानी से कर सकते हैं। माना हम उन वर्षों के आंकड़ों की रिपोर्ट तैयार करना चाहते हैं जिनकी पैदावार 1000 क्विंटल से अधिक और प्रतिफल भी रु 15000 से अधिक हो। ऊपर लिखे कोड में आंशिक परिवर्तन कर इस उद्देश्य को प्राप्त किया जा सकता है। कोड निम्न प्रकार है:-

```
proc print data=repo1 noobs split='/' width=uniform;
label yld=' YIELD (in ha)'
ret='RETURN (in Rs)'
pc='PRODUCTION COST (in Rs)'
profit='Profit Before *Taxes';
var year yld ret pc;
where yld>1000 or ret>15000;
```



```
sum yld ret pc profit;
```

```
title 'Report for Yield above 1000 and Return Rs  
15,000;
```

```
run;
```

इस कोड में width='uniform' परिणाम रिपोर्ट में सभी स्तंभों को बराबर चौड़ाई में विभक्त करता है और split='/' स्तम्भ शीर्षकों (columnheading) को दो या अधिक पंक्तियों में विभक्त करेगा। तुलना करने की शर्त को हमने WHERE कथन का प्रयोग कर, शर्त (condition) को परिभाषित कर दिया है। कोड को प्रस्तुत करने पर निम्न परिणाम प्राप्त होता है :

Report for Yield above 1000 Q and Return Rs 15,000				
PRODUCTION				
Year	YIELD (in hA)	RETURN (in Rs)	COST (in Rs)	Profit Before Taxes
1993	1098	16020	14600	1420
1994	1043	15645	15100	545
	2141	31665	29700	1965

PROC CHART PROC PLOT द्वारा साधारण रिपोर्ट (Simple Report by Using PROC Chart and PROC Plot)

जब आंकड़ों को, छोटी गणनाओं के साथ रिपोर्ट में प्रस्तुत करते हैं तो इसका एक अहम हिस्सा आंकड़ों का चार्ट या प्लॉट भी सम्मिलित किया जाना आवश्यक होता है। चार्ट और प्लॉट तैयार करने के लिए PROC CHART और PROC PLOT कथनों का प्रयोग अत्यंत आसानी के साथ कर सकते हैं। यदि हम उपरोक्त आंकड़ों की रिपोर्ट के साथ लाभ और पैदावार के आंकड़ों का चार्ट और प्लॉट भी तैयार करते हैं तो इसके लिए निम्न कोड लिखते हैं:-

```
proc chart data=repo1;
```

```
title 'Graphical Report on Yield (Q) Profit (Rs);
```

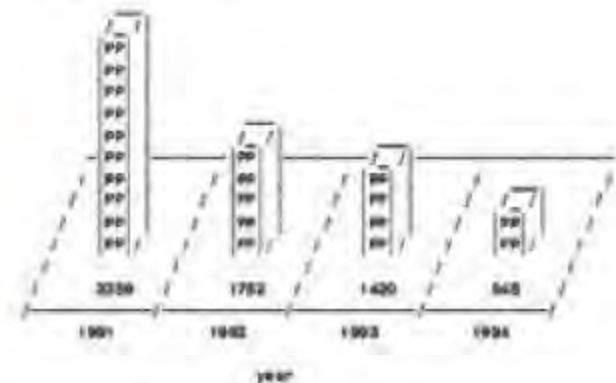
```
block year / sumvar=profit discrete symbol='P';
```

```
block year / sumvar=yld discrete symbol='Y';  
run;
```

प्रोग्राम को प्रस्तुत करने पर दो अलग अलग चार्ट तैयार हो जाते हैं। प्रोग्राम में हमने एक block में लाभ लिया जिसे चिन्ह 'P' से प्रदर्शित किया है और दूसरे block में पैदावार को लिया है जिसे 'Y' से प्रदर्शित किया है।

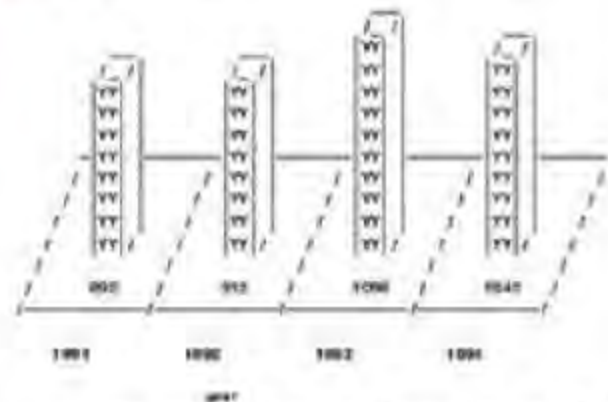
Graphical Report on Yield (Q) Profit (Rs)

Sum of Profit by year



Graphical Report on Yield (Q) Profit (Rs)

Sum of yld by year

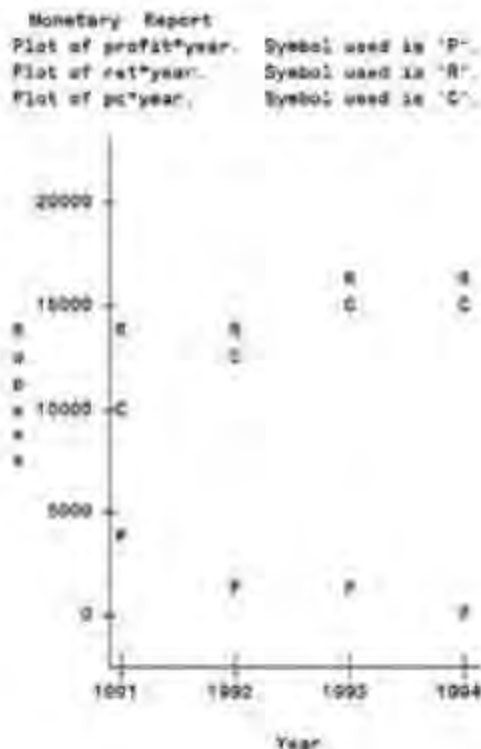


अब हम लाभ, प्रतिफल व पैदावार लागत, तीनों को एक ही अक्ष (axis) पर प्लॉट करते हैं क्योंकि इन तीनों का मात्रक (unit) रुपये है जिसे वर्ष के विरुद्ध प्लॉट किया जा सकता है। हम x अक्ष पर वर्ष और y अक्ष पर रुपये लेते हैं। परिणाम प्राप्ति के लिए प्रोग्राम निम्न प्रकार लिखते हैं :-



```
proc plot data=repo1;
title 'Monetary Report';
label profit='Rupees'
year='Year';
plot profit*year='P'
ret*year='R'
pc*year='C' / overlay vpos=20 hpos=32;
run;
```

यहाँ पर हमने, वर्ष के विरुद्ध लाभ प्रतिफल और पैदावार लागत को किया है इसलिए plot कथन में तीनों को इस प्रकार परिभाषित किया है कि x अक्ष पर वर्ष और y पर रूपये मात्रक रहे। प्रोग्राम तैयार करते समय y अक्ष के चर को * से पहले लिखना आवश्यक है। vpos=20 और hpos=32, ग्राफ में लम्ब और क्षैतिज रेखाएँ हैं जो ग्राफ के सेल बनती हैं, यह प्लॉट के बाह्य आयाम की बरकरार रखते हुये प्लॉट स्तम्भ में प्रदर्शित किए जा रहे अक्षरों P, R और C के आकार को सेल के आकार के अनुसार परिवर्तित कर देगा, ताकि वह सेल में ठीक प्रकार बैठ जाय। परिणाम निम्न प्रकार है:-



PROC REPORT द्वारा आंकड़ों की प्रस्तुति व सांख्यिकीय विश्लेषण

सारांश रिपोर्ट तैयार करने के PROC REPORT एक बहुत ही उपयोगी यंत्र है जोकि न केवल आंकड़ों की प्रस्तुति को सुंदर ढंग से प्रस्तुत करने में सहायक है बल्कि आधारभूत सांख्यिकी विश्लेषण गणना भी आसानी से करता है। उदाहरण के साथ प्रोग्राम दर्शाने से पूर्व यह आवश्यक होगा कि हम PROC REPORT में लिखे जाने वाले कथनों के बारे में आवश्यक जानकारी प्राप्त करले कि किस प्रकार कथन को ज्यादा उपयोगी बनाया जा सकता है।

DEFINE कथन का प्रयोग

DEFINE कथन यह वर्णित (specifies) करता कि रिपोर्ट के स्तम्भ (column) को कैसे गुणों (attributes) के साथ प्रदर्शित किया जाये। प्रत्येक चर के लिए एक DEFINE कथन COLUMN कथन के अन्दर उसके गुणों के साथ परिभाषित करके प्रयोग किया जाता है ताकि रिपोर्ट हमारे उद्देश्य के अनुसार तैयार हो सके। DEFINE कथन में प्रयोग होने वाले कुछ गुण इस प्रकार हैं:

- **Usage:** यह निदृष्ट करता है कि चर का प्रयोग कैसे करे। इसके लिए वैध विकल्प ACROSS, ANALYSIS, COMPUTED; DISPLAY, GROUP, ORDER है।
- **Attributes:** इसमें चरों के गुण FORMAT=, WIDTH= और SPACING= सम्मिलित है।
- **Options:** इसमें स्वरूपण (formatting) विकल्प DESCENDING, NOPRINT, NOZERO AND PAGE सम्मिलित है।
- **Justification:** इसका प्रयोग रिपोर्ट के स्तम्भ के औचित्य (Justification) जैसे CENTER, LEFT, RIGHT के लिए किया जाता है।
- **Column Heading:** इसका प्रयोग स्तम्भ शीर्षक (column heading) या चर के नाम के लेबल के लिए किया जाता है।

इन गुणों को लिखने के लिए सामान्य वाक्य



विन्यास निम्न प्रकार है जिसका आगे उदाहरण के साथ प्रयोग करके स्पष्ट किया गया है।

```
PROC REPORT DATA = <Dataset Name> NOWD;
```

```
COLUMN Var1 Var2 ----Var (n)
```

```
DEFINE variable/<usage> <attribute> <options>
```

```
<Justification> <column&heading>;
```

```
RUN;
```

PROC REPORT कथन में WINDOW या NOWINDOW का प्रयोग उपयोगकर्ता पर निर्भर करता है। व्यतिक्रम व्यवस्था (default setting) द्वारा WINDOW का प्रयोग ही होता है जिसकारण रिपोर्ट

विंडो में ही खुलती है। यह व्यवस्था हमारे द्वारा रिपोर्ट में बार-बार किये जा रहे परिवर्तनों को तुरंत उसी विंडो में देखने में हमें सक्षम बना देती है। यदि हम, हमारे द्वारा किये जा रहे परिवर्तनों को तुरंत उसी विंडो की बजाय एस ए एस आउटपुट विंडो में देखना चाहते हैं तो NOWINDOW कथन का प्रयोग कर सकते हैं।

PROC REPORT के कथनों को ठीक प्रकार समझने के लिए उचित होगा की हम एक उदाहरण के साथ स्पष्ट किया जाय। सर्वप्रथम हम एक एस ए एस डाटा सेट सर्जित करते हैं। माना हमारे पास एक परिक्षण के निम्न आंकड़े उपलब्ध हैं जिस पर हम विभिन्न प्रकार से रिपोर्ट तैयार करना चाहते हैं।

Centre	Year	Season	Crop	Variety	K Calories	GYield	SYield	Return
Varansi	2010	Kh	Rice	Swarna	19584	56.60	1340	50940
Varansi	2010	Kh	Rice	Swarna	13632	39.40	1480	35460
Varansi	2011	Kh	Rice	Swarna	15916	46.00	1248	41400
Varansi	2011	Kh	Rice	Swarna	16193	46.80	1350	42120
Varansi	2012	Kh	Rice	Swarna	17715	51.20	1090	46080
SK Nagar	2010	Kh	Pearmillet	GHB	20976	0.99	876	641
SKNagar	2010	Kh	Pearmillet	GHB	19555	5.85	967	3801
SK Nagar	2010	Kh	Pearmillet	GHB	21056	6.92	698	4496
SK Nagar	2011	Kh	Pearmillet	GHB	20761	10.50	655	6825
SK Nagar	2011	Kh	Pearmillet	GHB	19867	12.57	589	8170
Junagarh	2010	Kh	Clusterbean	GG1	18888	5.56	769	1667
Junagarh	2010	Kh	Pearlmillet	GHB	20889	23.65	978	12180
Junagarh	2011	Kh	Greengram	GM	17836	9.08	589	13528
Junagarh	2011	Kh	Cotton	Hy8	5021	14.44	700	28311
Junagarh	2012	Kh	Clusterbean	GG1	11890	3.97	239	1190
Varansi	2010	Rb	Wheat	HUW	17890	33.40	980	28390
Varansi	2010	Rb	Wheat	HUW	18564	24.60	990	20910
Varansi	2011	Rb	Wheat	HUW	19234	24.60	1000	20910
Varansi	2012	Rb	Wheat	HUW	20761	29.40	888	24990
Varansi	2012	Rb	Wheat	HUW	20789	33.00	967	28050



इस डाटा का एस ए एस डाटा सेट तैयार करने के लिए एस ए एस एडिटर में छोटा सा सामान्य प्रोग्राम लिखते हैं।

```
data crop_data;
input Centre $1-8 Year $10-13 Season $15-16 Crop $18-28 Variety $30-35
      KCalories   GYield   SYield   Return;
cards;
Varansi 2010 Kh Rice           Swarna 19584 56.60 1340 50940
Varansi 2010 Kh Rice           Swarna 13632 39.40 1480 35460
Varansi 2011 Kh Rice           Swarna 15916 46.00 1248 41400
Varansi 2011 Kh Rice           Swarna 16193 46.80 1350 42120
Varansi 2012 Kh Rice           Swarna 17715 51.20 1090 46080
SKNagar 2010 Kh Pearmillet    GHB    20976 0.99  876   641
SKNagar 2010 Kh Pearmillet    GHB    19555 5.85  967   3801
SKNagar 2010 Kh Pearmillet    GHB    21056 6.92  698   4496
SKNagar 2011 Kh Pearmillet    GHB    20761 10.50 655   6825
SKNagar 2011 Kh Pearmillet    GHB    19867 12.57 589   8170
Junagarh 2010 Kh Clusterbean  GG1    18888 5.56  769   1667
Junagarh 2010 Kh Pearlmillet  GHB    20889 23.65 978   12180
Junagarh 2011 Kh Greengram    GM     17836 9.08  589   13528
Junagarh 2011 Kh Cotton       Hy8    5021  14.44 700   28311
Junagarh 2012 Kh Clusterbean  GG1    11890 3.97  239   1190
Varansi 2010 Rb Wheat         HUW    17890 33.40 980   28390
Varansi 2010 Rb Wheat         HUW    18564 24.60 990   20910
Varansi 2011 Rb Wheat         HUW    19234 24.60 1000  20910
Varansi 2012 Rb Wheat         HUW    20761 29.40 888   24990
Varansi 2012 Rb Wheat         HUW    20789 33.00 967   28050
;
```

यदि हम इस डाटा सेट को प्रस्तुत (Submit) करे तो CROP_DATA के नाम से एसएसएस डाटा सेट बन जायेगा। आंकड़ों के ऊपर INPUT कथन में कुछ चरों का मान अक्षर (Character) है उनके साथ एक रिक्त स्थान के बाद डालर(\$) का चिह्न लगा है शेष बिना डालर चिह्न वाले चरों का मान संख्या है।

रिपोर्टिंग चरो का चयन

अपनी रिपोर्ट में यह निदृष्ट करने के लिए कि हम कौन-कौन से चरों को सम्मिलित करना चाहते हैं COLUMN कथन का प्रयोग करते हैं। यह कथन हमें उन्ही चरों की रिपोर्ट देगा जिन्हें परिभाषित किया जायेगा। यहाँ पर हम centre, year, season व crop चरों (variables) का चयन करते हैं जिन्हे COLUMN कथन मे परिभाषित करेंगे।

```
proc report data=crop -data NOWD;
```

```
column centre year season crop;
```

```
run;
```

अब हम इस कोड को प्रस्तुत (submit) कर देते हैं जिससे हमें निम्न परिणाम प्राप्त हुआ।

Centre	Year	Season	Crop
Varansi	2010	Kh	Rice
Varansi	2010	Kh	Rice
Varansi	2011	Kh	Rice
Varansi	2011	Kh	Rice
Varansi	2012	Kh	Rice
SKNagar	2010	Kh	Pearmillet
SKNagar	2010	Kh	Pearmillet
SKNagar	2010	Kh	Pearmillet
SKNagar	2011	Kh	Pearmillet
SKNagar	2011	Kh	Pearmillet



Centre	Year	Season	Crop
Junagarh	2010	Kh	Clusterbean
Junagarh	2010	Kh	Pearlmillet
Junagarh	2011	Kh	Greengram
Junagarh	2011	Kh	Cotton
Junagarh	2012	Kh	Clusterbean
Varansi	2010	Rb	Wheat
Varansi	2010	Rb	Wheat
Varansi	2011	Rb	Wheat
Varansi	2012	Rb	Wheat
Varansi	2012	Rb	Wheat

रिपोर्ट में पंक्तियों को समूहित करना व क्रम बदलना

रिपोर्ट में सारांश के उद्देश्य से पंक्तियों का एकत्रीकरण करने के लिए GROUP का प्रयोग किया जाता है। ORDER एक संकेत शब्द (keyword) है, इसका प्रयोग चरों के मान को रिपोर्ट में वर्णानुक्रमक (alphabetic) क्रम में प्रस्तुत करने के लिए किया जाता है। GROUP व ORDER की भूमिका स्पष्ट करने के लिए सबसे पहले एक एसएएस कोड तैयार करते हैं।

```
proc report data = crop_data NOWD MISSING;
column year centre crop GYield SYield Return ;
define year/group order;
define centre/group;
define crop/group order;
define GYield /display format = 6-2 "Grain Yield" ;
define SYield / display format = 6-2 "Straw Yield";
define Return / display format = 6-2 "Return (Rs/
ha)";
```

run;

इस एसएएस कोड के प्रथम पंक्ति में NOWD का प्रयोग किया है जो हमें, हमारे द्वारा कोडों में किए जाने वाले परिवर्तनों को परिणाम विंडो में दिखाएगा। इसी पंक्ति में MISSING भी लिखा है जो कि आंकड़ों गुम मान (missing value) का ध्यान रखेगा। COLUMN कथन में उन सभी चरों के नाम परिभाषित कर दिये हैं जिन्हें हम रिपोर्ट में सम्मिलित करना चाहते हैं।

DEFINE कथन सदैव COLUMN कथन के बाद आयेगा जिसमें हम प्रत्येक चर को अलग गुण के साथ उसी प्रकार परिभाषित करेंगे जैसा कि हम अपनी रिपोर्ट में चाहते हैं। चर year के बाद परोक्ष चिह्न (oblique sign) के साथ group व order दोनों परिभाषित किए हैं जो कि रिपोर्ट में प्रत्येक वर्ष का समूह बनाएगा और प्रत्येक समूह को एक के बाद एक क्रम से प्रदर्शित करेगा। पुनः centre चर का समूह बनेगा, अर्थात् प्रत्येक वर्ष के अंदर प्रत्येक केंद्र का समूह रिपोर्ट में प्रदर्शित करेगा इसीप्रकार हर crop चर जोकि फसल का नाम है, का समूह बनाकर क्रम से रिपोर्ट में प्रस्तुत होगा। अंत में GYield चर जोकि अनाज उत्पादन है को आठ अंकों (दशमलव के दो अंकों के साथ) प्रस्तुत करेगा तथा रिपोर्ट में चर GYield के स्थान पर "Grain Yield" प्रदर्शित करेगा।

अब हम इस कोड को प्रस्तुत करते हैं ताकि हमें इसका परिणाम प्राप्त हो सके। परिणाम इस प्रकार सारणी में प्राप्त हुआ।

Year	Centre	Crop	Grain Yield	Straw Yield	Return (Rs ha)	
2010	Junagarh	Clusterbean	5.56	769.00	1667.0	
		Pearlmillet	23.65	978.00	12180	
	SKNagar	Pearlmillet	0.99	876.00	641.00	
			5.85	967.00	3801.0	
	Varansi	Rice		6.92	698.00	4496.0
				56.60	1340.0	50940
Wheat			39.40	1480.0	35460	
			33.40	980.00	28390	
2011	Junagarh	Cotton	14.44	700.00	28311	
		Greengram	9.08	589.00	13528	
	SKNagar	Pearlmillet	10.50	655.00	6825.0	
			12.57	589.00	8170.0	
	Varansi	Rice	46.00	1248.0	41400	
			46.80	1350.0	42120	
2012	Junagarh	Wheat		24.60	1000.0	20910
				3.97	239.00	1190.0
	Varansi	Rice	51.20	1090.0	46080	
		Wheat	29.40	888.00	24990	
			33.00	967.00	28	



परिणाम सारणी रिपोर्ट में स्पष्ट है कि सबसे पहले year को समूहित किया है, year के अंदर Centre और पुनः इसके अंदर Crop को वर्णाक्रम में समूहित किया है।

उप-योग और कुल योग की गणना

हम अपनी रिपोर्ट में उप-योग (sub total) व कुल योग (grand total) भी प्रदर्शित कर सकते हैं परंतु इसके लिए हमारे पास उचित उद्देश्य होना आवश्यक है। हम एक रिपोर्ट तैयार करते हैं जिसमें वाराणसी केंद्र पर चावल के दाना उत्पादन (Grain Yield), पुआल (Straw) और प्रतिफल (Return) का वर्षवार योग और कुल योग प्रदर्शित हो। अब उद्देश्य की प्राप्ति के लिए एस ए एस कोड तैयार करते हैं। इसके लिए ऊपर लिखे कोड में कुछ जरूरी परिवर्तन करके आसानी से किया जा सकता है।

```
proc report data = crop_data NOWD missing ;
where centre = 'Varansi' and crop = 'Rice' ;
column year centre crop GYield SYield Return ;
define year/ order;
define rep/group;
define trt/order;
define centre/group;
define GYield/ "Wheat Yield (Q/ha)" format = 6-2;
define SYield/ "Straw (Q/ha)" format = 6-2;
define Return/ "Wheat Return (Rs/ha)" format = 6-2;
break after year/ summarize;
rbreak after/summarize;
run;
```

हमने यहाँ पर, where कथन के साथ centre और crop के नाम परिभाषित कर दिये हैं जिससे इसी crop से संबंधित आंकड़ों का रिपोर्ट में आंकलन होगा break after कथन तथा summarize गुण प्रत्येक वर्ष के अंत में उस वर्ष का योग की गणना करके रिपोर्ट में प्रदर्शित करेगा। अंत में सभी वर्षों में, केंद्र की परिभाषित की गयी फसल का कुलयोग प्रदर्शित करेगा। कोड को प्रस्तुत करने पर निम्न परिणाम प्राप्त हुआ।

Year	Centre	Crop	Wheat Yield (Q ha)	Straw (Q ha)	Wheat Return (Rs ha)
2010	Varansi	Rice	56.60	1340.0	50940
		Rice	39.40	1480.0	35460
2010			96.00	2820.0	86400
2011	Varansi	Rice	46.00	1248.0	41400
		Rice	46.80	1350.0	42120
2011			92.80	2598.0	83520
2012	Varansi	Rice	51.20	1090.0	46080
2012			51.20	1090.0	46080
			240.00	6508.0	216000

नये चर की गणना

एस ए एस के सभी PROC में गणना करने के लिए DATA स्टेप को प्रस्तुत (run) करना आवश्यक है परंतु PROC REPORT में नए चर की गणना सीधे की जा सकती जो इस PROC को एस ए एस के अंदर अनोखा बना देता है। माना, हम उपरोक्त आंकड़ों का कुल उत्पाद (Total Production) व शुद्ध प्रतिफल (Net Return) की गणना कर, दो नए चर अपनी रिपोर्ट में सर्जित करना चाहते हैं। कुल उत्पाद की गणना करने के लिए हम अनाज उत्पादन और पुआल उत्पादन को जोड़ देते हैं और शुद्ध प्रतिफल प्राप्त करने के लिए प्रतिफल में से परिवर्तनीय लागत (Variable Cost) को घटा देते हैं। माना, यहाँ पर गेहूँ (Wheat) उत्पादन पर परिवर्तनीय लागत रु 10,000 है। इस उद्देश्य की प्राप्ति के लिए निम्न कोड लिखते हैं—

```
proc report data = crop_data NOWD missing ;
where Centre = 'Varansi' and Crop = 'Wheat' ;
column Year Centre Crop GYield SYield TYield Return NR;
define Year/order;
define Centre/group;
define Crop/order width = 5 center;
define GYield/ "Grain Yield" format = 6-2;
define SYield/ "Straw Yield" format = 6-2;
define TYield/computed "Total Yield" format = 6-2;
define NR/computed "Net Return" format = 6-2;
compute TYield;
```



```

TYield = GYield-sum + SYield-sum;
endcomp;
compute NR;
NR = Return-sum & 10000 ;
endcomp;
title "Total Yield and Net Returns of Varanasi Centre"
;
break after Crop / summarize ;
run;

```

इस प्रोग्राम को प्रस्तुत करने पर निम्न परिणाम प्राप्त होगा

यहाँ पर हमने KCalories चर की सांख्यिकी गणना की जिसे column कथन में परिभाषित किया है फिर कोमा लगाकर कोष्ठक में सांख्यिकी गणना के लिए कुंजी शब्दों लिख दिये हैं। प्रोग्राम प्रस्तुत करने पर हमें निम्न परिणाम प्राप्त हो जाता है।

Calories						
Centre	N	Min	Max	Sum	Mean	Std
Varansi	5.00	13632.00	19584.00	83040.00	16608.00	2213.01

Total yield and Net Return of Varanasi Centre

Year	Centre	Crop	Grain Yield	Straw Yield	Total Yield	Return	Net Return
2010	Varansi	Wheat	33.40	980.00	1013.4	28390	18390
			24.60	990.00	1014.6	20910	10910
2011	Varansi	Wheat	24.60	1000.0	1024.6	20910	10910
2012	Varansi	Wheat	29.40	888.00	917.40	24990	14990
			33.00	967.00	1000.0	28050	18050

विश्लेषण चर की सांख्यिकीय

PROC REPORT से हम मूलभूत सांख्यिकीय विश्लेषण सीधे आसानी से प्राप्त कर सकते हैं। हम Column कथन में कोष्ठक के अंदर जो भी मूलभूत सांख्यिकीय प्राप्त करना चाहते हैं उसे परिभाषित कर देते हैं। निम्न प्रोग्राम द्वारा हम कुल चर संख्या, न्यूनतम मान, अधिकतम मान, औसत व मानक विचलन, मूलभूत सांख्यिकीय की गणना करते हैं।

```

proc report data = Crop_data NOWD missing;
where crop = 'Rice' ;
column Centre KCalories, (N Min Max Sum Mean
Std);
define Year/group;
define Centre/group;
define KCalories/ "Calories" format=6-2;
run;

```

PROC TABULATE द्वारा साधारण रिपोर्ट

PROC Tabulate बिलकुल उसी तरह सांख्यिकीय गणना करता है जैसे की हम PROC MEANS या PROC FREQ द्वारा करते हैं। PROC TABULATION सांख्यिकीय आंकड़ों को प्रस्तुत करने एवं विश्लेषण करने में अत्यंत सहायक है। इसके द्वारा हम नए चर को उत्पन्न कर पुनः विश्लेषण के नए डाटा सेट के अंदर डाल सकते हैं। उपरोक्त डाटा सेट का उपयोग कर, हम अलग-अलग तीन सारणियाँ तैयार करते हैं जिसमें हमें इच्छित सूचनाएँ प्राप्त हो सकेंगी। PROC TABULATION का उपयोग करते हुये हम पहली 2X2 की सारणी में फसल के अनाज उपज, पुआल उपज, कलोरी व प्रतिफल के औसत आंकड़ों को प्रस्तुत करेंगे। दूसरी सारणी में हम प्रत्येक केंद्र तथा प्रत्येक ऋतु के अनाज उपज व पुआल उपज, चरों के औसत आंकड़े प्राप्त करेंगे तथा तीसरी सारणी में वर्षवार व प्रत्येक केंद्र के अनाज उपज, पुआल उपज व कलोरी चरों के औसत आंकड़ें प्रस्तुत करेंगे। इन औसत आंकड़ों को



प्राप्त करने लिए एस ए एस में एक छोटा सा प्रोग्राम तैयार करते हैं जो निम्न प्रकार है :-

```
proc tabulate data=nars;
class Year centre crop season;
var GYield Return Kcal SYield;
title 'Cross Tabular Summary Report';
table crop, mean*(GYield SYield Kcal Return);
table centre (Season), mean**(GYield Return);
table Year (Centre), mean*(KCal GYield SYield);
run;
```

प्रोग्राम को प्रस्तुत करने पर हमें निम्न तीन सारणीयन प्राप्त होंगी :-

Cross Tabular Summary Report

Crop	Mean			
	GYield	SYield	KCal	Return
Clusterbean	4.77	504.00	15389.00	8930.00
Greengram	9.08	589.00	17836.00	13528.00
Pearmillet	7.37	757.00	20443.00	36951.60
Rice	48.00	1301.60	16608.00	43200.00
Wheat	27.00	939.00	19662.50	22950.00

Centre	Mean	
	GYield	Return
Junagarh	6.20	10462.67
SKNagar	7.37	36951.60
Varansi	42.00	37414.29
Season	22.73	33242.00
Kh		
Rb	27.00	22950.00

Year	KCal	Mean	
		GYield	SYield
2010	18893.57	19.99	1017.14
2011	18114.60	24.99	886.20
2012	16788.67	28.19	739.00
Centre	16204.67	6.20	532.33
Junagarh			
SKNagar	20443.00	7.37	757.00
Varansi	17480.71	42.00	1198.00

सारांश

बायोलोजिकल अन्वेषकों को प्रायः विस्तृत सांख्यिकीय विश्लेषण की आवश्यकता नहीं होती परन्तु संबन्धित बायोलोजिकल अन्वेषण की रिपोर्ट तैयार करने के लिए आधारभूत सांख्यिकीय विश्लेषण और आंकड़ों की प्रस्तुति की आवश्यकता होती है। PROC PRINT, PROC TABULATION एवं PROC REPORT अन्वेषकों को अपने अनुसंधान कार्य को सुंदर ढंग से प्रस्तुत करने समर्थ बना देता है। अन्वेषक न केवल संबन्धित अनुसंधान कार्य के आंकड़ों को प्रस्तुत करने में सक्षम होता है बल्कि आधारभूत सांख्यिकीय विश्लेषण व चार्ट और ग्राफ को भी आसानी से अपने अनुसंधान कार्य में प्रस्तुत कर सकता है।

संदर्भ

[http:// support-sas-com](http://support-sas-com)

<http://analytics-ncsu-edu/sesug>

<http://www2-sas-com/proceedings/sugi2>

सी 4.5 एल्गोरिथ्म का प्रयोग करते हुए निर्णय-वृक्ष वर्गीकरण एवं विजुअलाइजेशन के लिए एक ऑनलाइन सॉफ्टवेयर (ओडीटीसी)

शशि दहिया, सुवजीत दास, अंशु भारद्वाज एवं ऊषा जैन

सार

डाटा माइनिंग में वर्गीकरण एक महत्वपूर्ण एवं व्यापक रूप से करने वाला कार्य है। यह एक प्रागुक्त मॉडलिंग कार्य है जिसे व्याख्यात्मक चरों (explanatory variables) के फलन के रूप में लक्ष्य चर के लिए एक मॉडल बनाने के रूप में परिभाषित किया गया है। वर्गीकरण की अनेक सुस्थापित तकनीकें उपलब्ध हैं, जिसमें (निर्णय-वृक्ष), मशीन लर्निंग डोमेन से एक महत्वपूर्ण एवं लोकप्रिय तकनीक है।

निर्णय-वृक्ष, एक निर्णय सपोर्ट टूल है जो वृक्ष की तरह के ग्राफ अथवा निर्णय लेने वाले मॉडल एवं उनके संभावित परिणामों का प्रयोग करता है जिसमें चान्स इवेंट आउटकम्स, संसाधन लागत एवं उपयोगिता शामिल हैं। सी4.5 डाटासेटों के वर्गीकरण में प्रयुक्त एक सुज्ञात निर्णय वृक्ष एल्गोरिथ्म है। सी4.5 एल्गोरिथ्म क्विनलेन्स द्वारा बनाई गयी आईडी3 एल्गोरिथ्म का विस्तार है जो कि निर्णय-वृक्ष के वर्गीकरण हेतु उपयोग में लाई जाती है। यह निर्णय-वृक्ष को इन्ड्यूज करता है और डाटासेटों से नियम जनरेट करता है, जिनमें स्पष्ट तथा/अथवा संख्यात्मक विशेषताएं हो सकती हैं। सी4.5 डाटासेटों को वर्गीकृत करने के साथ-साथ उपयोगी नियमों को बनाने में अच्छा प्रदर्शन करता है।

इस शोध-पत्र में, सी4.5 एल्गोरिथ्म का प्रयोग करते हुए नियम बनाने एवं निर्णय वृक्ष इन्डक्शन के लिए एक वेब आधारित सॉफ्टवेयर पर चर्चा की गयी है। वृक्ष संरचना के रूप में विजुअलाइजेशन, उत्पन्न नियमों की समझ को बढ़ाता है। इस सॉफ्टवेयर में आंकड़ों में लुप्त मानों को इम्प्यूट करने की सुविधा

उपलब्ध है। इनपुट डाटा स्पष्ट एवं संख्यात्मक, दोनों प्रकृति का हो सकता है। यह सॉफ्टवेयर TXT, XLS एवं CSV डाटा फाइल फॉरमेट्स को इम्पोर्ट कर सकता है। सॉफ्टवेयर के विकास की प्रक्रिया के लिए संवर्धित वाटरफॉल मॉडल का प्रयोग किया गया है। यह सॉफ्टवेयर शिक्षाविदों, शोधकर्ताओं एवं डाटा माइनिंग, कृषि एवं अन्य क्षेत्रों में कार्य कर रहे छात्रों के लिए जहां बड़ी मात्रा में आंकड़े जनरेट होते हैं, बहुत उपयोगी होगा।

प्रस्तावना

आंकड़ों के विश्लेषण एवं व्याख्या के लिए ज्ञान में आंकड़ों का ट्रांसफारमेशन अधिकांशतः मैनुअल विधियों का प्रयोग करते हुए किया गया है। आंकड़ों का आकार बढ़ा होने की स्थिति में इससे डाटाबेसों के पैटर्न एक्स्ट्रेक्शन की प्रक्रिया बहुत खर्चीली, धीमी गति की एवं अत्यधिक वस्तुनिष्ठ होती है। डाटाबेसों में ज्ञान डिस्कवरी (कंडीडी) आंकड़ों की सतत समझ एवं अर्थ प्राप्त करने के साथ-साथ नया ज्ञान प्राप्त करने के उद्देश्य से आंकड़ों के बृहत् समूह से ज्ञान एक्स्ट्रेक्शन की एक प्रक्रिया है। यह प्रक्रिया बहुत जटिल है क्योंकि इसमें एक प्रौद्योगिकी निहित है जो सॉफ्टवेयर के गणितीय एवं तकनीकी मॉडलों के समूह से गठित है और जो आंकड़ों की प्रवृत्ति एवं उनकी अनियमितताओं का पता लगाने के लिए प्रयोग होती है। कंडीडी को आंकड़ों की वैधता, नोवल एवं संभावित उपयोगी पैटर्न की खोज करने की प्रक्रिया के रूप में परिभाषित किया गया है। कंडीडी प्रक्रिया इन्टरेक्टिव एवं पुनरावृत्त है जिसमें निम्नलिखित बुनियादी चरण शामिल हैं: आंकड़ों का चयन, आंकड़ों का ट्रांसफॉर्मेशन, डाटा माइनिंग,

प्रवृत्ति मूल्यांकन एवं ज्ञान प्रतिनिधित्व। डाटा माइनिंग केडीडी प्रक्रिया का एक अनिवार्य चरण है, जिसमें डाटा प्रवृत्ति प्राप्त करने के लिए इन्टेलीजेन्ट विधियों का प्रयोग किया जाता है। डाटा माइनिंग को "अन्तर्निहित, पूर्व अज्ञात, एवं डाटा से संभावित उपयोगी जानकारी के नॉनट्रीवीयल निकास" के रूप में परिभाषित किया जा सकता है (हान, 2006)। व्यवहार में डाटा माइनिंग के दो "उच्च-स्तरीय" मुख्य उद्देश्य प्रागुक्ति एवं व्याख्या हैं। डाटा माइनिंग के कार्य को प्रायः दो प्रमुख श्रेणियों में विभाजित किया जाता है : भविष्यवाणी कार्य एवं व्याख्यात्मक कार्य। प्रागुक्ति एवं व्याख्या के लक्ष्य को प्राप्त करने के लिए डाटा माइनिंग से सम्बन्धित विभिन्न कार्य किए जाते हैं, जैसेकि वर्गीकरण एसोसिएशन रूल माइनिंग, क्लस्टरिंग, आउटलायर डिटेक्शन, डिपेन्डेन्सी मॉडलिंग इत्यादि। वर्गीकरण, जो अनेक पूर्वनिर्धारित श्रेणियों में से किसी एक को उद्देश्य सौंपने का कार्य है, एक व्यापक समस्या है जिसमें अनेक विविध अनुप्रयोग शामिल हैं। यह डाटा माइनिंग का सबसे अधिक महत्वपूर्ण एवं व्यापक रूप से किये जाने वाला कार्य है। वर्गीकरण के लिए निर्णय-वृक्ष (डीटी) एक लोकप्रिय एवं व्यापक रूप से प्रयुक्त मशीन लर्निंग तकनीक है। निर्णय वृक्ष एक फ्लोचार्ट की तरह वृक्ष की संरचना है, जहां प्रत्येक आन्तरिक नोड (नॉन-लीफ नोड) एक एट्रीब्यूट पर एक परीक्षण को दर्शाती है और प्रत्येक शाखा, परीक्षण के परिणाम को दर्शाती है तथा प्रत्येक लीफ नोड (अथवा टर्मिनल नोड) में एक क्लास लेबल है। कुछ निर्णय-वृक्ष एल्गोरिथ्म ID3, C4.5, NBTree, SimpleCart, REPtree, BFtree एवं अन्य हैं। एक निर्णय-वृक्ष की रूट नोड का चयन करने के लिए तीन एट्रीब्यूट चयन मापों का प्रयोग किया जाता है। ये माप हैं: सूचना लाभ, लाभ अनुपात एवं गिनी सूचकांक। निर्णय वृक्ष की रूट नोड का चयन करने के लिए प्रत्येक एट्रीब्यूट में निहित सूचना की गणना के लिए विभिन्न निर्णय वृक्ष एल्गोरिथ्म में विभिन्न मापों का प्रयोग होता है।

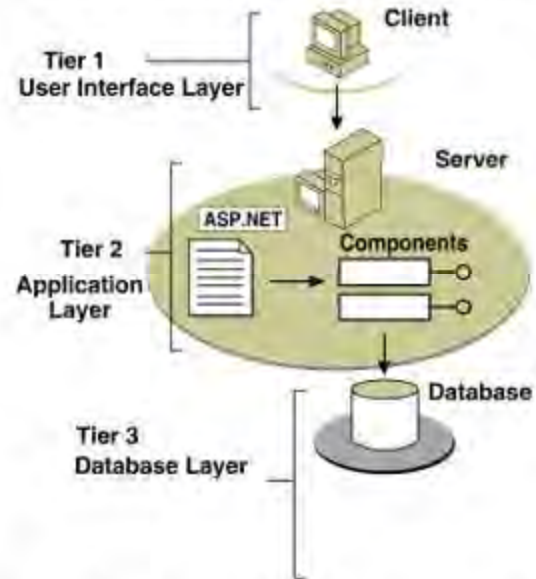
सॉफ्टवेयर आर्किटेक्चर

संशोधित वाटरफॉल विकास मॉडल का प्रयोग करते हुए (सी4.5) एल्गोरिथ्म के प्रयोग द्वारा ऑनलाइन

निर्णय वृक्ष वर्गीकरण (ओडीटीसी) विकसित किया गया है। यह एक लेयर्ड संरचना के रूप में क्रियान्वित है जिसमें प्रत्येक लेयर एक अलग कार्यक्षमता दर्शाती है। इस तंत्र के विकास के लिए मानक 3-टायर क्लाइन्ट सर्वर आर्किटेक्चर का प्रयोग किया गया है। इस तंत्र के कार्य को निम्नलिखित तीन लेयर्स में विभाजित किया गया है :

1. क्लाइन्ट साइड इन्टरफेस लेयर (सी.एस.आई.एल)
2. सर्वर साइड एप्लीकेशन लेयर (एसएसएएल)
3. डाटा बेस लेयर (डीबीएल)

आरेख 1 : ओडीटीसी की 3-टायर आर्किटेक्चर



सी.एस.आई.एल यह सुनिश्चित करता है कि तन्त्र प्रयोक्ता फ्रैण्डली हो। सी.एस.आई.एल में प्रयोक्ता से सूचना स्वीकार करने के लिए डाटा प्रपत्र होते हैं तथा जावा स्क्रिप्ट का प्रयोग करते हुए यह इन फार्म्स का वैधीकरण करता है। यह सबसे अधिक महत्वपूर्ण लेयर है क्योंकि यह प्रयोक्ता एवं सॉफ्टवेयर के बीच एक इन्टरफेस के रूप में कार्य करता है। यह एचटीएमएल, कासकेडिंग स्टाइल शीट (सीएसएल) एम्बेडेड जावा स्क्रिप्ट का प्रयोग करते हुए क्रियान्वित किया गया है। एसएसएएल में वेबसर्वर पर एप्लीकेशन लॉजिक शामिल है जो क्लाइन्ट साइड एवं डाटाबेस लेयर्स के बीच संचार स्थापित कर सकता है। एसएसएएल, ASP.NET YseodZ पर विजुअल C# का प्रयोग करते



सी 4.5 एल्गोरिथम का प्रयोग करते हुए निर्णय-वृक्ष वर्गीकरण एवं विजुअलाइजेशन के लिए एक ऑनलाइन सॉफ्टवेयर (ओडीटीसी)

हुए क्रियान्वित है। डायनमिक वेब पेजेज सृजित करने के लिए ASP.NET एक शक्तिशाली एवं फ्लेक्सीबल प्रौद्योगिकी है। यह दो प्रमुख माइक्रोसॉफ्ट प्रौद्योगिकियों, एक्टिव सर्वर पेजेज (एएसपी) एवं .NET फ्रेमवर्क, का मिश्रण है (मैकडॉनाल्ड, 2002)।

डीबीएल. एसक्यूएल सर्वर 2008 का प्रयोग करते हुए क्रियान्वित है। एप्लीकेशन के साथ डाटाबेस की कनेक्टिविटी के लिए ADO.NET का प्रयोग किया गया है। ADO.NET डाटा एक्सेस प्रौद्योगिकी है जो .NET फ्रेमवर्क में बनी है (एस्पोजिटो, 2005)। सॉफ्टवेयर वेबसाइट पर पंजीकरण करने वाले प्रत्येक प्रयोक्ता के लॉगइन एवं अन्य सम्बन्धित विवरण रिकार्ड करने तथा इनपुट डाटा से जनरेट हुए निर्णय नियम को स्टोर करने के लिए डाटाबेस का प्रयोग किया गया है।

सॉफ्टवेयर विवरण

ओडीटीसी एक ऑनलाइन सॉफ्टवेयर है, जो प्रयोक्ताओं को किसी भी समय मुफ्त सुलभ है। ओडीटीसी के लिए सुरक्षा सुनिश्चित की जाती है क्योंकि केवल प्राधिकृत प्रयोक्ता ही सॉफ्टवेयर को एक्सेस कर सकते हैं। यह एक निर्णय नियम जनरेशन सॉफ्टवेयर है जिसका उद्देश्य मुख्यतः विशेषज्ञ तंत्रों के विकास को सुलभ बनाना है जिन्हें नियमों के रूप में ज्ञान की आवश्यकता होती है। ओडीटीसी शोधकर्ताओं को अपनी उपलब्धियों से नियम तैयार करने में मदद करता है जो निर्णय निर्माण एवं विश्लेषणात्मक गतिविधियों को सुलभ बना सकता है। यह निर्णय वृक्ष के रूप में जनरेट किये गए नियमों के लिए विजुअलाइजेशन सुविधा उपलब्ध कराता है जो जनरेटेड नियमों की समझ को बढ़ाता है। इन नियमों का निर्यात भी किया जा सकता है तथा इन्हें एक्सेल फॉरमेट में सेव किया जा सकता है। ओडीटीसी निर्णय वृक्ष की प्रिण्टिंग की सुविधा भी उपलब्ध कराता है।

प्रयोक्ता प्रबन्धन

ब्राउजर से इस सॉफ्टवेयर को ब्राउज करने के पश्चात प्रयोक्ता को एक नया अकाउन्ट बनाना होगा अथवा वर्तमान अकाउन्ट का प्रयोग करते हुए लॉगइन

करना होगा। यदि प्रयोक्ता अपना पासवर्ड भूल जाता है, तो ओडीटीसी पासवर्ड रिसेट करने का विकल्प उपलब्ध कराता है (आरेख 2)।

आरेख 2 : नया अकाउन्ट बनाना अथवा वर्तमान अकाउन्ट का प्रयोग करते हुए लॉगइन करना



इनपुट डाटा प्रबन्धन

सफलतापूर्वक लॉगइन करने के पश्चात, प्रयोक्ता तीन फाइल फॉरमेट में उपलब्ध अपने आंकड़ों का वर्गीकृत कर सकता है। इस तंत्र द्वारा समर्थित ये तीन फाइल फॉरमेट्स हैं – एक्सेल, सीएसवी एवं टेक्स्ट फॉरमेट (आरेख 3)

आरेख 3 : फाइल अपलोड के लिए तीन विकल्प



लुप्त मान हैंडलिंग

ओडीटीसी सॉफ्टवेयर अपलोड किए गए डाटासेट अथवा फाइल में लुप्त मानों का पता लगाने व उन्हें इम्प्यूट करने की सुविधा उपलब्ध कराता है। यह लुप्त

सेल्स को उपयुक्त मानों से भी भरता है। डाटासेट में सभी लुप्त मान प्रश्न चिन्ह (“?”) सहित लाल रंग के सेल से प्रतिस्थापित किये जाएंगे।

समूह एट्रीब्यूट्स के मामले में, लुप्त मान सम्बन्धित एट्रीब्यूट के सभी विशिष्ट मानों के MODE (उच्चतम बारम्बारता) द्वारा इम्प्यूट किये जाएंगे। संख्यात्मक एट्रीब्यूट्स के मामले में, लुप्त मान सम्बन्धी एट्रीब्यूट के उन सभी मानों के अर्थमैटिक मीन (ARITHMETIC MEAN) द्वारा इम्प्यूट किये जाएंगे जो उसी वर्ग से सम्बन्धित हैं जिससे लुप्त मान वाले दृष्टांत (instance) सम्बन्धित हैं। यह तंत्र पंक्तियों, एट्रीब्यूट्स तथा सेल्स की कुल संख्या को लुप्त मानों के साथ प्रदर्शित करता है (आरेख 4)।

आरेख 4 : ओडीटीसी द्वारा हैण्डल किये गए लुप्त मान

outlook	temp	humidity	wind	play	sea
sunny	hot	high	weak	no	
sunny	hot	high	strong	no	
overcast	hot	high	weak	no	
overcast	medium	high	weak	no	
rainy	cool	low	weak	no	
rainy	cool	low	strong	no	
overcast	cool	low	strong	no	
sunny	hot	high	weak	no	
sunny	cool	low	weak	no	
rainy	medium	low	no		
sunny	medium	low	strong	no	
overcast	medium	high	strong	no	
overcast	hot	low	weak	no	
rainy	medium	high	strong	no	

डाटासेट पार्टिशन एवं एट्रीब्यूट चयन

वर्गीकरण के लिए ओडीटीसी को एक ही डाटासेट से डाटा के दो सेट्स की आवश्यकता होती है। अतः, इनपुट डाटा को दो भागों में नामतः प्रशिक्षण डाटा तथा परीक्षण डाटा में विभाजित करने की आवश्यकता है (शैम्यूलि इत्यादि, 2006)। ओडीटीसी प्रयोक्ता की वरीयता के अनुसार डाटासेट को प्रशिक्षण एवं परीक्षण डाटासेटों में विभाजित करने की सुविधा उपलब्ध कराता

है। परीक्षण डाटा का चयन प्रयोक्ता द्वारा किया जा सकता है तथा वह इसका चयन दृष्टांत (instance) की संख्या के रूप में कुल आंकड़ों के कम से कम 6 प्रतिशत तथा अधिक से अधिक 40 प्रतिशत का चयन करके कर सकता है। उदाहरणतया; यदि प्रयोक्ता द्वारा 150 इन्स्टेन्सेस का डाटासेट अपलोड किया गया है तो वह 8 (150 का 5 प्रतिशत) से 60 (150 का 40 प्रतिशत) के बीच इन्स्टेन्सेस की संख्या परीक्षण डाटा के रूप में चुन सकता है। डाटासेट का शेष भाग प्रशिक्षण डाटासेट के रूप में माना जाएगा। फिर भी, प्रयोक्ता परीक्षण डाटासेट के लिए पूर्ण डाटासेट से अपनी इच्छानुसार इन्स्टेन्सेस का भी चयन कर सकता है। परीक्षण डाटासेट के लिए इन्स्टेन्सेस की बांछित संख्या का चयन स्क्रीन के शीर्ष पर उपलब्ध स्लाइडिंग बार का प्रयोग करके किया जा सकता है। (आरेख-5)।

वर्गीकरण के लिए परीक्षण डाटा इन्स्टेन्सेस तथा



आरेख 5 : ओडीटीसी द्वारा एट्रीब्यूट चयन

आश्रित एवं स्वतंत्र एट्रीब्यूट का चयन पूरा करने के पश्चात प्रयोक्ता को आगे जाने के लिए ‘प्रोसीड’ बटन क्लिक करना होगा। ‘प्रोसीड’ बटन को क्लिक करने के पश्चात एक पोपअप बॉक्स दिखाई देगा जिसमें प्रयोक्ता अपने चयन के अनुसार प्रशिक्षण एवं परीक्षण



सी 4.5 एल्गोरिथम का प्रयोग करते हुए निर्णय-वृक्ष वर्गीकरण एवं विजुअलाइजेशन के लिए एक ऑनलाइन सॉफ्टवेयर (ओडीटीसी)

Data for Classification				
Training Dataset				
sl	sw	pl	pw	class
5.1	5.9	1.4	0.2	Iris-setosa
4.9	5	1.4	0.2	Iris-setosa
4.7	3.2	1.3	0.2	Iris-setosa
4.6	3.1	1.5	0.2	Iris-setosa
5	5.0	1.4	0.2	Iris-setosa
5.4	5.9	1.7	0.4	Iris-setosa
4.6	3.4	1.4	0.3	Iris-setosa
5	3.4	1.5	0.2	Iris-setosa
4.4	2.9	1.4	0.2	Iris-setosa
49	3.1	1.5	0.1	Iris-setosa

Data for Classification				
Test Dataset				
sl	sw	pl	pw	class
6.4	2.8	5.6	2.1	Iris-virginica
7.2	3	5.8	1.6	Iris-virginica
7.4	2.8	6.1	1.9	Iris-virginica
7.9	3.8	6.4	2	Iris-virginica
6.4	2.8	5.6	2.2	Iris-virginica
6.3	2.8	5.1	1.5	Iris-virginica
6.1	2.6	5.6	1.4	Iris-virginica
7.7	3	6.1	2.3	Iris-virginica
6.3	3.4	5.6	2.4	Iris-virginica
6.4	3.1	5.5	1.8	Iris-virginica

आरेख 6 : ओडीटीसी में प्रशिक्षण एवं परीक्षण डेटासेट का विभाजन

डाटा सेट में विभाजित, वर्गीकरण के लिए तैयार किए गए डेटासेट को देख सकता है। (आरेख 6)

रूल जनरेशन

जब प्रयोक्ता प्रशिक्षण डेटासेट एवं परीक्षण डेटासेट का विभाजन देखता है तो उसके पास चयन प्रक्रियाओं में किसी भी प्रकार का परिवर्तन करने के लिए विकल्प होता है। यदि प्रयोक्ता किसी प्रकार का परिवर्तन नहीं करना चाहता है तो रूल जनरेट करने के लिए उसे "Confirm" बटन पर क्लिक करना

Rule_sl	IF THEN RULES	Rows Covered
1.	IF pl >= 2.45 THEN Iris-setosa	47
2.	IF pl > 2.45 AND IF pw <= 1.65 AND IF pl <= 4.95 THEN Iris-versicolor	43
3.	IF pl > 2.45 AND IF pw <= 1.65 AND IF pl <= 4.95 THEN Iris-virginica	1
4.	IF pl > 2.45 AND IF pw <= 1.65 AND IF pl > 5.05 THEN Iris-virginica	21
5.	IF pl > 2.45 AND IF pw <= 1.65 AND IF pl <= 5.05 AND IF sw <= 2.85 THEN Iris-virginica	5
6.	IF pl > 2.45 AND IF pw <= 1.65 AND IF pl <= 5.05 AND IF sw > 2.85 AND IF sl <= 5.00 THEN Iris-versicolor	1
7.	IF pl > 2.45 AND IF pw <= 1.65 AND IF pl <= 5.05 AND IF sw > 2.85 AND IF sl > 5.95 AND IF sl <= 6.15 THEN Iris-virginica	1
8.	IF pl > 2.45 AND IF pw <= 1.65 AND IF pl <= 5.05 AND IF sw > 2.85 AND IF sl > 5.95 AND IF sl <= 6.15 THEN Iris-versicolor	1

Export If Then Rule to Excel

SNo.	sl	sw	pl	pw	class	Predicated_Class	Match Found
1.	6.4	2.8	5.6	2.1	Iris-virginica	Iris-virginica	Y
2.	7.2	3	5.8	1.6	Iris-virginica	Iris-virginica	Y
3.	7.4	2.8	6.1	1.9	Iris-virginica	Iris-virginica	Y
4.	7.9	3.8	6.4	2	Iris-virginica	Iris-virginica	Y
5.	6.4	2.8	5.6	2.2	Iris-virginica	Iris-virginica	Y
6.	6.3	2.8	5.1	1.5	Iris-virginica	Iris-virginica	Y
7.	6.1	2.6	5.6	1.4	Iris-virginica	Iris-virginica	Y
8.	7.7	3	6.1	2.3	Iris-virginica	Iris-virginica	Y

Export Test Result to Excel

Test accuracy is : 96.67%

Export Confusion Matrix to Excel

Show Tree View

Actual/Predicted	Iris-setosa	Iris-versicolor	Iris-virginica	Precision
Iris-setosa	5	0	0	1
Iris-versicolor	0	4	1	0.8
Iris-virginica	0	0	22	1
Recall	1	1	0.96	0

Export Confusion Matrix to Excel

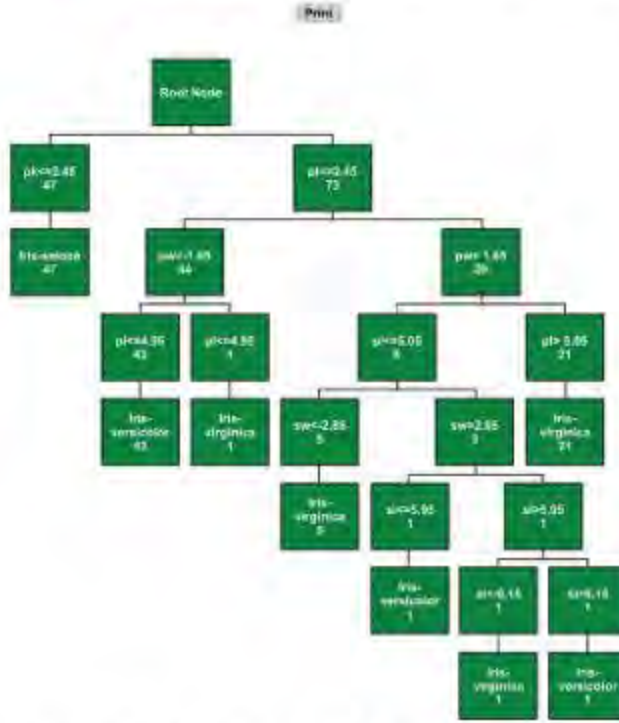
Show Tree View

आरेख 7 : इरिस डेटासेट के लिए ओडीटीसी द्वारा तैयार किया गया परिणाम पृष्ठ

होगा। रूल के साथ-साथ प्रयोक्ता परीक्षण डेटा सेट के "रूल कवरेज" (एक विशेष रूल द्वारा कवर किए गए इन्स्टेंसेस की संख्या), "Test accuracy", "confusion matrix" की सूचना प्राप्त कर सकता है। प्रयोक्ता को इन सभी को एक्सेल फॉरमेट में निर्यात करने की सुविधा प्राप्त है (आरेख 7)।

निर्णय वृक्ष विजुअलाइजेशन

ओडीटीसी के इस पृष्ठ के नीचेले भाग में प्रयोक्ता निर्णय वृक्ष के विजुअलाइजेशन के लिए "Show Tree View" नामक मॉड्यूल देखा सकता है जो जनरेट किए गए रूल्स की समझ को बढ़ाता है। "Show Tree View" बटन पर क्लिक करने से प्रयोक्ता सॉफ्टवेयर द्वारा तैयार किए गए निर्णय वृक्ष को देख सकता है तथा इस निर्णय वृक्ष का प्रिंट आउट लेने का भी विकल्प दिखाई देगा। (आरेख 8)



आरेख 8 : इरिस डाटासेट के लिए ओडीटीसी द्वारा तैयार किये गए निर्णय-वृक्ष का चित्र

परिकल्पित (मौसम सम्बन्धी डाटासेट) एवं सैम्पल डाटा मॉड्यूल में उपलब्ध कराये गए बैच मार्क (इरिस डाटासेट) डाटासेटों का प्रयोग करते हुए इस सॉफ्टवेयर को वैधिकृत किया गया।

निष्कर्ष

ओडीटीसी से प्राप्त निष्कर्ष तुलना योग्य थे तथा कुछ मामलों में, कुछ बैचमार्क एवं कुछ अन्य डाटासेटों के साथ जांच करने पर बेहतर पाए गए। ओडीटीसी, निर्णय रूल्स के जनरेशन तथा उनके मूल्यांकन मापों की एक ऑनलाइन सुविधा है। यह प्रयोक्ता फ्रेंडली सॉफ्टवेयर है तथा इसमें कम्प्यूटर प्रोग्रामिंग की विशेषज्ञता की आवश्यकता नहीं है। प्रयोक्ता

पंजीकरण करवा सकता है, लागइन कर सकता है, रूल जनरेट कर सकता है, परिणामों एवं निर्णय वृक्ष का विजुअलाइजेशन करने के साथ-साथ परिणामों को एक्सेल में एक्सपोर्ट भी कर सकता है। प्रयोक्ता को पास परिणामों को प्रिंट करने का भी विकल्प होता है। सी 4.5 एल्गोरिथ्म का प्रयोग करते हुए बड़े डाटासेटों में छिपे पैटर्न का पता लगाने के लिए इस तंत्र का प्रयोग किया जा सकता है। इस प्रकार, इसे ज्ञान जनरेशन सॉफ्टवेयर के रूप में भी प्रयोग किया जा सकता है जो निर्णय रूल्स एवं निर्णय वृक्ष के रूप में एक विशेष डाटासेट के सम्बन्ध में जानकारी एक्सट्रेक्ट करने के साथ-साथ उपलब्ध भी करा सकता है। जनरेट की गयी जानकारी का प्रयोग शोधकर्ताओं, शिक्षाविदों, छात्रों एवं विकासकों द्वारा किया जा सकता है।

संदर्भ

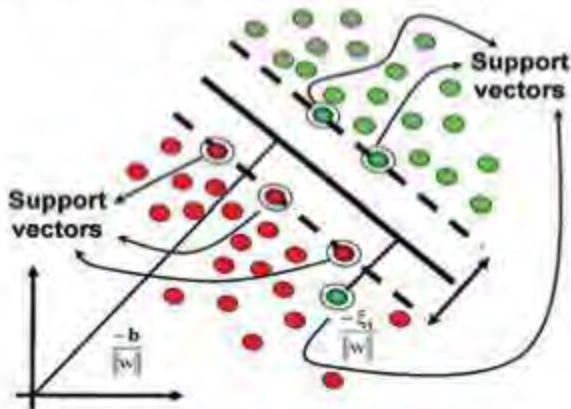
1. एस्पोजिटो, डी. (2005) – प्रोग्रामिंग माइक्रोसॉफ्ट एएसपी.नेट 2.0 – डब्ल्यू.पी. पब्लिशर्स ऐण्ड डिस्ट्रीब्यूटर्स प्राइवेट लिमिटेड, बंगलौर, इण्डिया।
2. हान, जे. एवं कैम्बर (2006) – डाटा माइनिंग-कॉन्सेप्ट्स ऐण्ड टेक्नीक्स – मॉर्गन कॉफमैन पब्लिशर्स।
3. मैकडोनाल्ड, एम. (2002) – एएसपी.नेट: दि कम्प्लीट रेफ्रेन्स। टाटा मैकग्रॉ हिल पब्लिशिंग कम्पनी लिमिटेड, नई दिल्ली, इण्डिया।
4. शैम्यूलि, जी., पटेल, एन.आर. एवं ब्रूस, पी.सी. (2006) – डाटा माइनिंग फॉर बिजनेस इण्टेलीजेन्स - कॉन्सेप्ट्स, टेक्नीक्स, ऐण्ड एप्लीकेशन्स इन माइक्रोसॉफ्ट ऑफिस एक्सेल विद एक्स.एल. माइनर जॉहन विल्ले ऐण्ड संस, आईएनसी. न्यू जर्सी।

सपोर्ट वेक्टर समाश्रयण से काल श्रृंखला आँकड़ों का पूर्वानुमान: भारतीय तिलहन उत्पादन डेटा में अनुप्रयोग

संतोष राठोड़, के. एन. सिंह, प्रकाश कुमार, राजीव रंजन कुमार, सुशील कुमार सरकार, बिशाल गुरुंग, मृन्मय राय एवं अचल लामा

परिचय

सपोर्ट वेक्टर मशीन (एस.वी.एम.) एक लर्निंग तकनीक है जिसे मूलतः रैखिक वर्गीकरण समस्याओं के लिए विकसित किया गया था। बाद में वर्ष 1997 में, वैपनिक द्वारा लॉस फंक्शन लागू कर समाश्रयण समस्याओं के लिए सपोर्ट वेक्टर मशीन विकसित की गयी तथा यह अरैखिक समाश्रयण आकलन समस्याओं तक बढ़ा दी गयी है। ऐसी समस्याओं की मॉडलिंग को अरैखिक सपोर्ट वेक्टर समाश्रयण मॉडल (वैपनिक, 2001) कहा जाता है। एस.वी.एम. का आधारभूत सिद्धांत मूल इनपुट काल श्रृंखलाओं को हाई डाइमेंशनल फीचर स्पेस में ट्रांसफार्म करना तथा नई फीचर स्पेस में समाश्रयण मॉडल का निर्माण करना है। कृषि, उद्योग, स्टॉक मार्केट मूल्य पूर्वानुमान इत्यादि अनेक क्षेत्रों में काल श्रृंखला प्रागुक्ति में अरैखिक सपोर्ट वेक्टर समाश्रयण का व्यापक उपयोग किया गया है (हाँग, इत्यादि 2006, कॉंग, इत्यादि 2016, कुमार एवं प्रज्ञेय, 2015)। निम्नलिखित चित्र में सपोर्ट वेक्टर अवधारणा को दर्शाया गया है (आकृति 1)।



आकृति 1: सपोर्ट वेक्टरों के योजनाबद्ध वर्णन

सपोर्ट वेक्टर समाश्रयण

मान लिया जाए $Z = \{x_i, y_i\}_{i=1}^N$ डेटा सेट वेक्टर है जहाँ $x_i \in R^n$ है इसमें वेक्टर ऑफ इनपुट और वेक्टर ऑफ आउटपुट $x_i \in R^n$ दोनों हैं, N डेटा सेट की साइज है, तो नान लीनियर सपोर्ट वेक्टर रिग्रेशन का सामान्य रूप इस प्रकार होगा।

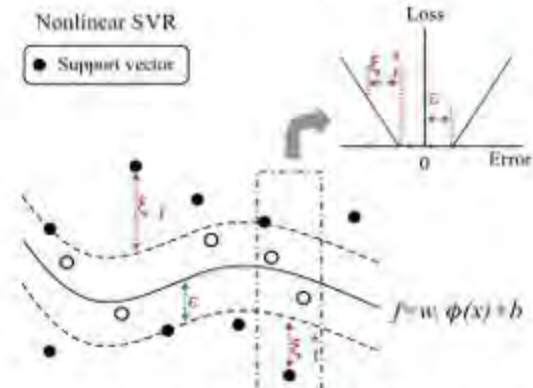
$$f(x) = W^T \phi(x) + b$$

जहाँ $\phi(\cdot): R^n \rightarrow R^{nh}$ एक नान लिनियर मैपिंग फंक्शन है, $w \in R^{nh}$ वेट वेक्टर है, b एक वायस है। कोफिशिएंट w और b को निम्न रिस्क फंक्शन को मिनिमाइज्ड कर प्राप्त कर सकते हैं:

$$R(\theta) = \frac{1}{2} \|w\|^2 + C \left[\frac{1}{N} \sum_{i=1}^N L_\epsilon(y_i, f(x_i)) \right]$$

समीकरण (2) में दो घटक शामिल है, पहला रेगुलेराइज्ड टर्म है i.e. $\frac{1}{2} \|w\|^2$ और $\frac{1}{N} \sum_{i=1}^N L_\epsilon(y_i, f(x_i))$ इम्पिरिकल एरर टर्म है, जिसे निम्न वैपनिक ϵ -इन्टेनसिव लॉस फंक्शन से आकलन कर सकते हैं

$$L_\epsilon(y_i, f(x_i)) = f(x) = \begin{cases} |y_i - f(x_i)| - \epsilon; & |y_i - f(x_i)| \geq \epsilon \\ 0 & |y_i - f(x_i)| < \epsilon \end{cases}$$



आकृति 2: अरैखिक SVM मॉडल सेटअप के तहत वैपनिक-इन्संसिटिव लॉस फंक्शन और सटीकता ट्यूबका योजनाबद्ध वर्णन



जहाँ y_t मूल वेल्यु है तथा $f(x_t)$ एस्टीमेटेड वेल्यु है। अरैखिक सपोर्ट वेक्टर रिग्रेशन का अंतिम रूप इस प्रकार होता है:-

$$f(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(x_i, x_j) + b, t = 1, 2, \dots, N \quad (4)$$

जहाँ α_i एवं α_i^* को लैग्रेज मल्टीप्लायर कहते हैं। अरैखिक एस.वी.आर. मॉडल सेटअप के तहत वैपनिक-इन्संसिटिव लॉस फंक्शन और सटीकता ट्यूब को निम्नलिखित आकृति में दर्शाया गया है (आकृति 2).

अरैखिक सपोर्ट वेक्टर रिग्रेशन मॉडलिंग में ऑप्टीमल हाइपर प्लेन का चुनाव करना एक महत्वपूर्ण कदम होता है। अरैखिक सपोर्ट वेक्टर रिग्रेशन का प्रदर्शन कर्नल फलन (तालिका 1) और हाइपर पैरामीटर पर निर्भर करता है।

ϵ के वेल्यु को ट्युब साइज कहते हैं। ϵ_1 और ϵ_2 ϵ $-i^*$ दो पोजीटिव स्लेक वेरियेबल हैं जो ϵ -Vigo के मूल

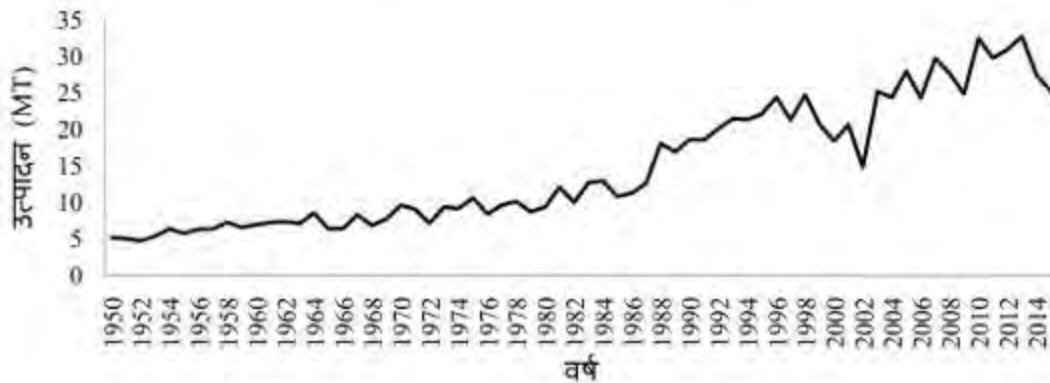
वेल्यु तथा बाउंड्री वेल्यु के बीच की दूरी को प्रदर्शित करता है। अगर डेटा ट्युब के अंदर पाया जाता है तो ये वेल्यु शून्य होता है। इन स्लेक वेरीएबल का प्रयोग सपोर्ट वेक्टर की संख्या ज्ञात करने में किया जाता है।

उदाहरण:

1950-51 से 2015-16 तक भारत का कुल तिलहन उत्पादन (MT) के वार्षिक आंकड़े भारतीय रिजर्व बैंक (RBI), भारत सरकार (आरबीआई सांख्यिकी, 2016) द्वारा जारी कृषि सांख्यिकी से एकत्र किए गए हैं। 1950-51 से 2010-11 के आंकड़ों का इस्तेमाल मॉडल संरचना के लिए और 2011-12 से 2015-16 के लिए मॉडल सत्यापन के लिए किया गया है। वर्णनात्मक सांख्यिकी और काल श्रृंखला की समीक्षा के तहत श्रृंखला की प्लॉट क्रमशः तालिका 2 और आकृति 3 में दी गई है।

तालिका 1: सपोर्ट वेक्टर मशीन समस्या में आमतौर पर उपयोग में आने वाले कर्नल फलन

क्र. संख्या	कर्नल के प्रकार	व्यंजक
1.	रेखक SVM	$K(x, x_i) = x_i^T x$
2.	Polynomial of degree d	$K(x, x_i) = (x_i^T x + k)^d$
3.	Radial Basis Function (RBF)	$K(x, x_i) = \exp\left\{-\frac{\ x-x_i\ ^2}{2\sigma^2}\right\}$ Equivalently $K(x, x_i) = \exp\{-\gamma\ x-x_i\ ^2\}$
4.	Multi-Layer Perceptron (MLP)	$K(x, x_i) = \tanh(k_1 x_i^T x + k_2)$



आकृति 3: भारत का तिलहन उत्पादन का काल श्रृंखला प्लाट



तालिका 2: तिलहन उत्पादन का वर्णनात्मक सांख्यिकी

Statistic	तिलहन उत्पादन	Statistic	तिलहन उत्पादन
Observation	66	Maximum	32.75
Mean	14.86	Standard Deviation	8.47
Median	11.05	Skewness	0.6
Mode	6.4	Kurtosis	-1.04
Minimum	4.73	Coefficient of Variation (%)	56.98

तिलहन उत्पादन काल श्रृंखला के लिए अरैखिक सपोर्ट वेक्टर रिग्रेशन (NLSVR) मॉडल निम्नलिखित पैरामीटर विनिर्देशों (तालिका 3) के साथ बनाया गया है। माना हुआ काल श्रृंखला के लिए क्रॉस सत्यापन भी किया गया है और निम्नतम क्रॉस सत्यापन त्रुटि 0.035 है। इसके अलावा प्रशिक्षण सेट और परीक्षण डेटा सेट में मॉडल का प्रदर्शन तालिका 5 और 6 में दिया गया है।

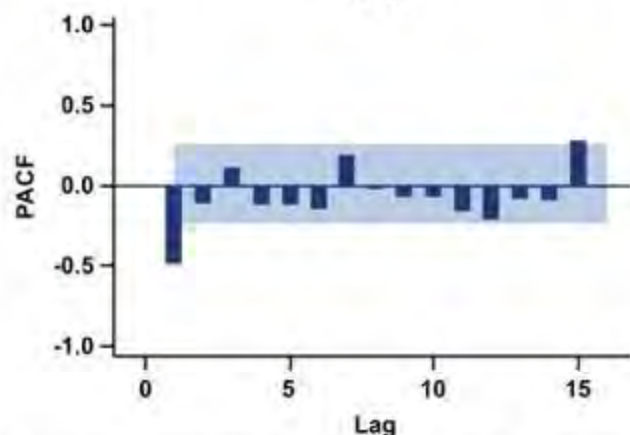
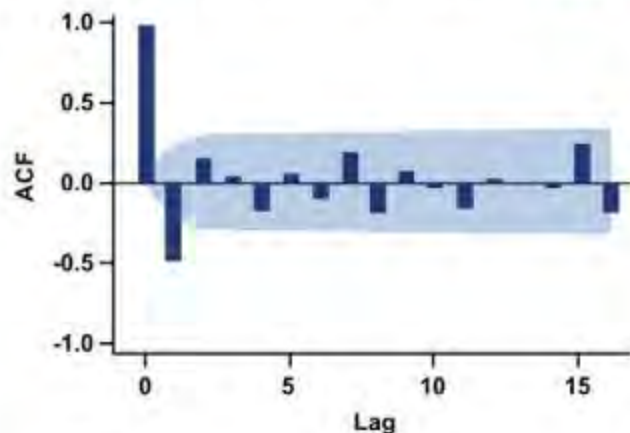
तालिका 3: तिलहन उत्पादन काल श्रृंखला के लिए एसवीआर का मॉडल विनिर्देश

Kernel function	No. of SVs	C γ ε	K fold cross validation (K)	Cross Validation Error
RBF	7	8.19 3.060.15	10	0.035

Univariate ARIMA मॉडल फिटिंग

भारत के तिलहन उत्पादन के लिए ARIMA मॉडल का निर्माण किया गया है। मूल काल-श्रृंखला गैर स्थिर था, इसलिए पहले श्रृंखला को अंतर निकल कर स्थिर टाइम सीरीज बनाया गया है (आकृति 4)।

पर्याप्त मॉडल अर्थात् ARIMA (1,1,0) की पहचान ऑटोकोरिलेशन और आंशिक ऑटोकोरिलेशन फंक्शन (ICF एवं PACF) प्लॉट्स (आकृति 4) के आधार पर की गई है। मँगो प्रोडक्शन टाइम सीरीज के ARIMA मॉडल से प्राप्त अवशिष्टों के लिए ऑटो सहसंबंध की जांच से पता चलता है कि कार्ई-स्क्वेयर का मान 0.45 के रूप में गैर-स्वायत्तता वाले अवशेष पाए गए



आकृति 4. भारत का तिलहन उत्पादन काल श्रृंखला का ACF एवं PACF

हैं। इसके अलावा प्रशिक्षण सेट और परीक्षण डेटा सेट में मॉडल का प्रदर्शन 5 और 6 तालिका में दिया गया है।

तालिका 4: तिलहन उत्पादन काल श्रृंखला के लिए मैक्सिमम लाइकलीहूड अनुमान विधि द्वारा ARIMA (1,1,0) का पैरामीटर अनुमान

Parameter	Estimate	Standard Error	t Value	Approx. Pr > t	Lag
MU	0.43	0.19	1.64	0.1012	0
AR1,1	-0.56	0.18	-4.39	<0.0001	1

तालिका 5: प्रशिक्षण डेटा सेट के लिए तिलहन उत्पादन काल श्रृंखला का मॉडल प्रदर्शन

Criteria	ARIMA	NLSVR
MSE	5.33	1.39
RMSE	2.32	1.18
MAPE	11.64	6.88



तालिका 6 : परीक्षण डेटा सेट के लिए तिलहन उत्पादन काल श्रृंखला का मॉडल प्रदर्शन

Year	Actual	Forecast ARIMA	NLSVR
2012	30.94	31.54	31.42
2013	32.75	30.72	33.64
2014	27.51	31.88	30.15
2015	25.30	31.90	26.77
Criteria	MSE	13.31	2.06
	RMSE	3.64	1.43
	MAPE	10.58	4.24

न्यूनतम मीन स्क्वेयर एरर (MSE) के आधार पर, रूट मीन स्क्वेयर एरर (RMSE) और मॉडलों का मीन एब्सोल्यूट पेरसेंटेज (MAPE) के मान दोनों प्रशिक्षण (तालिका 3) और परीक्षण (मान्यकरण) डेटा सेट (तालिका 4) दोनों के लिए प्राप्त किए गए हैं, जिससे यह अनुमान लगा सकता है कि NLSVR तकनीक ने ARIMA मॉडल पर बेहतर प्रदर्शन किया। हालांकि, भिन्नता का गुणांक (तालिका 1) बहुत अधिक है, फिर भी कृत्रिम बुद्धिमान तकनीक अर्थात् NLSVR बेहतर प्रदर्शन करता है। इसका कारण यह हो सकता है कि एकचरी ARIMA मॉडल से तुलना में अरैखिक मशीन लर्निंग तकनीक डेटा सेट में विषम प्रवृत्ति को कैच कर सकती है और बेहतर प्रदर्शन करती है।

निष्कर्ष

ARIMA मॉडल अरैखिक संरचनाओं वाले काल श्रृंखला के लिए हमेशा पर्याप्त नहीं होते। इस संदर्भ में, अरैखिक कृत्रिम लर्निंग की तकनीक जैसे की सपोर्ट वेक्टर मशीनों से पूर्वानुमान से कार्यक्षमता को बेहतर

बनाने का एक प्रभावी तरीका हो सकती है। इस काम में प्राप्त परिणामों के आधार पर, यह अनुमान लगाया सकता है कि समय-सीमा के मॉडलिंग और पूर्वानुमान के लिए अरैखिक वेक्टर रिग्रेशन तकनीक से पूर्वानुमान की सटीकता बढ़ाई जा सकती है। अरैखिक सपोर्ट वेक्टर रिग्रेशन मॉडल ने अन्य मॉडल के मुकाबले तिलहन उत्पादन के पूर्वानुमान के लिए बेहतर प्रदर्शन किया है। ऑटोरिग्रेसिव और मूवींग एवरेज का उपयोग भी मशीन लर्निंग तकनीकों में करके इस दृष्टिकोण को आगे बढ़ाया जा सकता है।

संदर्भ:

काँग, वाई., वेंग, जे. एवं ली एक्स (2016): ट्रैफिक फ्लो फोरकास्टिंग बाई ए लीस्ट स्क्वेयर्स सपोर्ट वेक्टर मशीन विद ए फ्रूट फलाई ऑप्टिमाइजेशन एल्गोरिथ्म, प्रोसिडिया इंजीनियरिंग, 137,59-68.

कोटर्स, सी. एवं वैपनिक, वी (1995): सपोर्ट वेक्टर नेटवर्क, मशीन लर्निंग, 20, 1-25.

हाँग, डब्ल्यू सी. एवं पाई, पी.एफ. (2006): प्रिडिक्टिंग इंजन रिलायबिलिटी बाई सपोर्ट वेक्टर मशीन्स . इण्टरनेशनल जर्नल ऑफ एंडवान्सड मैनुफैक्चरिंग टेक्नोलॉजी, 28, 154-161.

कुमार, टी.एल.एम. एवं प्रज्ञेय (2015): डवलपमेंट ऑफ हाइब्रिड मॉडल्स फॉर फोरकास्टिंग टाइम सीरिज डाटा यूजिंग नॉन-लीनियर एस.वी.आर. एनहेन्सड बाई पी.एस.ओ., जर्नल ऑफ स्टैटिस्टिकल थ्योरी ऐण्ड प्रैक्टिस, 9(4): 699-711.

वैपनिक, वी (2000): दि नेचर ऑफ स्टैटिस्टिकल लर्निंग थ्योरी, सेकेन्ड एडिशन, सिप्रिंगर-वर्ल्ड, न्यूयॉर्क

अनुपात का लघु क्षेत्र आकलन

सुशील कुमार, हुकुम चन्द्र, वन्दिता कुमारी, कौस्तव आदित्य एवं प्रदीप बसाक

सार

दोहरे विकल्प डाटा के लघु क्षेत्र आकलन में लिनियर मिक्स्ड मॉडल आधारित सामान्य लघु क्षेत्र आकलक का इस्तेमाल त्रुटिपूर्ण हो जाता है। अतः दोहरे विकल्प डाटा से लघु क्षेत्र अनुपात के आकलन हेतु लॉगिट फंक्शन युक्त जनरलाइज्ड लिनियर मिक्स्ड मॉडल आधारित एम्पिरिकल प्लग-इन प्रेडिक्टर (EPP) का इस्तेमाल किया जाता है। यहाँ EPP के इस्तेमाल के लिए पूर्ण आबादी का ईकाई स्तरीय सहायक चर डाटा की जरूरत होती है, जोकि सदा उपलब्ध नहीं होता है। इस स्थिति से निपटने के लिए हमने इकाई स्तर प्रतिदर्श सहायक चर डाटा आधारित लघु क्षेत्र अनुपात आकलक एवं इसके मीन स्क्यूरेड एरर (MSE) आकलक प्रस्तावित किया है। मॉटे कार्लो सिमुलेशन आधारित परिणाम दिखाता है कि प्रदित आकलक मौजूद आकलक से बेहतर प्रदर्शन करता है।

मुख्य शब्द: लॉगिट फंक्शन, जनरलाइज्ड लिनियर मिक्स्ड मॉडल, दोहरे विकल्प डाटा।

परिचय

लघु क्षेत्र आधारित छोटे नमूनों के उपयोग द्वारा विश्वसनीय सांख्यिकी की मांग ने लघु क्षेत्र आकलन के सैद्धांतिक एवं व्यावहारिक दृष्टिकोण के विकास को बढ़ावा दिया है। लघु क्षेत्र आकलन के पारंपरिक आकलनकर्ता (एस्टिमेटर) प्रायः क्षेत्र विशिष्ट नमूना के आकार बाध्यता के कारण अस्थिर होते हैं। इस परिप्रेक्ष्य में उचित लघु क्षेत्र मॉडल आधारित पद्धतियाँ "उधार शक्ति" (borrowing strength) के कार्यान्वयन द्वारा कुशल अनुमानकों के निर्माण की अनुमति देती हैं। लघु क्षेत्र मॉडल आधारित अनुमानक प्रायः स्पष्ट लिंकिंग मॉडल का उपयोग करते हैं। ये अनुमानक क्षेत्र विशिष्ट प्रभावों पर आधारित होते हैं जो क्षेत्र भिन्नता की व्याख्या

करते हैं। निरंतर प्रतिक्रिया वेरिएबल्स (continuous response variable) के लिए रेखीय मिश्रित मॉडल के तहत empirical best linear unbiased predictor (EBLUP) दृष्टिकोण का उपयोग बहुत ही आम है और छोटे क्षेत्र आकलन हेतु इसे कुशल माना जाता है, देखें Rao एवं Molina (2015)।

कई सर्वेक्षणों में दोहरे विकल्प (बाइनरी) डाटा अक्सर रुचिकर होते हैं, विशेषकर जब हमारा उद्देश्य सर्वेक्षण जनसंख्या की व्यवसाय समूह बनाना हो, जैसे लाभदायक बनाम गैर-लाभदायक व्यवसाय, अभिनव बनाम गैर-अभिनव व्यवसाय आदि। इसी तरह गरीबी और स्वास्थ्य संबंधी संकेतक चर आधारित सर्वेक्षण डाटा भी बाइनरी होता है। इस तरह के डाटा हेतु प्रायः logit लिंक फंक्शन (referred as logistic linear mixed model) युक्त सामान्यीकृत रेखिक मिश्रित मॉडल (जीएलएमएम) का उपयोग किया जाता है। सहायक चर की सूचना उपलब्धता के आधार पर, लघु क्षेत्र आकलन में प्रयुक्त मॉडल को प्रायः क्षेत्र स्तर या इकाई स्तर मॉडल में वर्गीकृत किया जाता है। यदि कुछ सहायक चर का इकाई स्तरीय माप सभी जनसंख्या इकाइयों के लिए सुलभ हो तो EPP का नियोजन इकाई स्तरीय जीएलएमएम के तहत किया जाता है। परंतु व्यावहारिक स्थिति में सहायक चर के इकाई स्तरीय माप हमेशा उपलब्ध नहीं होता है। नतीजतन, कई लघु क्षेत्र अनुप्रयोगों में, EPP अनुमानक का उपयोग नहीं किया जा सकता। यदि सहायक चर की सकल स्तरीय आबादी की जानकारी उपलब्ध हो तो क्षेत्र स्तरीय GLMM के तहत EPP का इस्तेमाल किया जा सकता है। यद्यपि सहायक सूचनाओं की उपलब्धता की कई स्थितियाँ हो सकती हैं, परंतु इस लेख में हमने ऐसी स्थिति को माना है जिसमें इकाई स्तरीय सहायक सूचना केवल नमूना डाटा के लिए उपलब्ध है और



गैर-नमूना इकाइयों का संबंधित इकाई स्तरीय सहायक चर की जानकारी उपलब्ध नहीं है। उदाहरण के लिए, भारत में राष्ट्रीय नमूना सर्वेक्षण कार्यालय (NSSO) द्वारा एकत्र किए गए घरेलू उपभोक्ता व्यय सर्वेक्षण डाटा और ऋण निवेश सर्वेक्षण डाटा। ऐसी स्थिति में लघु क्षेत्र आकलनकर्ता को प्रत्यक्ष सर्वेक्षण आधारित आकलन करना पड़ता है। अतः इस अवस्था में लघु क्षेत्र आकलन विधियों का पता लगाना दिलचस्प है। इस लेख में हमने, लॉगिट (Logit) लिंक फंक्शन युक्त जीएलएमएम के तहत लघु क्षेत्र अनुपात आकलन का प्रस्ताव किया है।

अनुपात का लघु क्षेत्र आकलन

माना कि D लघु क्षेत्रों में विभाजित U एक जनसंख्या है जहां U_i ($i=1, \dots, D$) का आकार N_i है और $\bigcup_{i=1}^D U_i = U$ एवं $\sum_{i=1}^D N_i = N$ है। आबादी U से ' n ' आकार का एक नमूना s चुना गया, जिसमें लघु क्षेत्र i से n_i आकार का उप-नमूना s_i चुना गया है, इस प्रकार से $\bigcup_{i=1}^D s_i = s$ और $\sum_{i=1}^D n_i = n$ है। यहाँ r_i क्षेत्र i के गैर-नमूना इकाइयों को निर्देशित करता है। y_{ij} लघु क्षेत्र i के j वां इकाई के माप को निर्देशित करता है। यह एक दोहरे विकल्पधारी चर है। यहाँ हमारा उद्देश्य लघु क्षेत्र के आबादी अनुपात का अनुमान लगाना है।

$$P_i = N_i^{-1} \sum_{j \in U_i} y_{ij} = N_i^{-1} \left\{ \sum_{j \in s_i} y_{ij} + \sum_{j \in U_i \setminus s_i} y_{ij} \right\}$$

यहाँ अनुपात का डिजाइन-आधारित प्रत्यक्ष अनुमानक (DIR) एवं विचरण क्रमशः

$$\hat{P}_i^{DIR} = \sum_{j \in s_i} w_{ij} y_{ij} \text{ और}$$

$$\text{var}(\hat{P}_i^{DIR}) = \sum_{j \in s_i} w_{ij} (w_{ij} - 1) (y_{ij} - \hat{P}_i^{DIR})^2$$

यहाँ $w_{ij} = w_{ij}^* / \sum_{j \in s_i} w_{ij}^*$ क्षेत्र i में इकाई j की सामान्यीकृत सर्वेक्षण वजन हैं, $\sum_{j \in s_i} w_{ij}^* = 1$ और w_{ij}^* क्षेत्र i में इकाई j की सर्वेक्षण भार है। विचरण अनुमानक, *Särndal et al. (1992), pp- 43, 185 और 391* से प्राप्त किया गया है जिसका सरल रूप $w_{ij} = 1/h_{ij}$, $h_{ij} = h_{ij}$ और $h_{ij,k} = h_{ij} h_{ik}$, $j \neq k$, जहाँ h_{ij} क्षेत्र i में इकाई j का नमूना में शामिल होने की प्रथम ऑर्डर संभाव्यता है और $h_{ij,k}$ क्षेत्र i में इकाई j और k का दूसरे

क्रम में शामिल होने की संभाव्यता है। साधारण यादृच्छिक नमूनाकरण विधि के तहत $w_{ij}^* = N_i n_i^{-1}$ है इसका प्रत्यक्ष अनुपात अनुमानक $\hat{P}_i^{DIR} = p_i = n_i^{-1} \sum_{j \in s_i} y_{ij}$ एवं विचरण $\text{var}(\hat{P}_i^{DIR}) = n_i^{-1} \hat{P}_i^{DIR} (1 - \hat{P}_i^{DIR})$ हैं। प्रत्यक्ष आकलन निष्पक्ष होता है, परंतु केवल क्षेत्र-विशिष्ट नमूना डाटा पर आधारित होने के कारण प्रत्यक्ष अनुमानक अस्थिर हो जाता है। नतीजतन, लघु क्षेत्र आकलन में प्रत्यक्ष अनुमानक को उत्साहित नहीं किया जाता है। इसके अलावा, उच्च नमूना परिवर्तनशीलता के वजह से इन अनुमानों का विश्वास अंतराल भी बहुत बड़ा हो जाता है। साथ ही साथ नमूना विहीन क्षेत्रों के लिए भी प्रत्यक्ष अनुमानक का उपयोग नहीं किया जा सकता।

इस संदर्भ में, मॉडल-आधारित लघु क्षेत्र आकलन (SAE) विधियां, जिसमें सांख्यिकीय मॉडल के माध्यम से 'शक्ति उधार लिया जाता है' का उपयोग विश्वसनीय छोटे क्षेत्र अनुमान के उत्पादन के लिए किया जा सकता है, (Rao and Molina (2015))। इन विधियों का उपयोग मॉडल आधारित अप्रत्यक्ष अनुमानकों के निर्माण हेतु किया जाता है। SAE विधियां स्पष्ट लिंकिंग मॉडल का उपयोग करती हैं जो मॉडल में शामिल सहायक चर द्वारा व्याख्यित क्षेत्र अंतरों से परे क्षेत्र-विशिष्ट प्रभावों की व्याख्या करता है।

लघु क्षेत्र अनुपात आकलन हेतु दोहरे विकल्प डाटा के लिए प्रायः लॉगिट लिंक फंक्शन युक्त जीएलएमएम आधारित EPP का इस्तेमाल किया जाता है, (Chandra et al., 2012)। माना कि \mathbf{x}_{ij} बाइनरी y_{ij} चर से संबंधित p इकाई स्तर मानों का वेक्टर है। π_{ij} क्षेत्र i में यूनिट j की मान 1 प्राप्त करने की संभाव्यता है। इसके लिए लॉगिस्टिक लिंक फंक्शन युक्त जीएलएमएम इस प्रकार से है

$$\text{logit}(\pi_{ij}) = \log \left[\frac{\pi_{ij}}{1 - \pi_{ij}} \right] = \eta_{ij} = \mathbf{x}_{ij}' \boldsymbol{\beta}_i + u_{ij}, \quad (1)$$

$$j = 1, \dots, N_i; i = 1, \dots, D$$

यहाँ $\boldsymbol{\beta}_i$ एक p फिक्स्ड इफैक्ट परमिटर मानों का वेक्टर है। u_{ij} क्षेत्र i का क्षेत्र विशिष्ट प्रभाव है जिसका वितरण $u_{ij} \sim N(0, \sigma^2)$ हैं। हमारा मानना है की u_{ij} 's आपस में स्वतंत्र है और $y_{ij} | u_{ij} \sim \text{Binomial}(1, \pi_{ij})$, जहाँ $E(y_{ij} | u_{ij}) = \pi_{ij}$ और $\text{Var}(y_{ij} | u_{ij}) = \sigma_{ij}^2 = \pi_{ij}(1 - \pi_{ij})$ ।



मॉडल (1) के अंतर्गत, $E(y_{ij} | u_i) = \pi_{ij} = \exp(\eta_{ij}) / \{1 + \exp(\eta_{ij})\}^{-1}$ इसमें $\eta_{ij} = x_{ij}^T \beta + u_i$ । यहाँ लघु क्षेत्र i के लिए लघु क्षेत्र अनुपात का EPP

$$\hat{P}_i^{EPP} = N_i^{-1} \left(\sum_{j \in S_i} y_{ij} + \sum_{j \in U_i} \hat{\mu}_{ij} \right) \quad (3)$$

जहाँ $\hat{\mu}_{ij} = \hat{E}(y_{ij} | u_i) = \exp(x_{ij}^T \hat{\beta} + \hat{u}_i) / \{1 + \exp(x_{ij}^T \hat{\beta} + \hat{u}_i)\}^{-1}$ । यहाँ $\hat{\beta}$ और \hat{u}_i क्रमशः मॉडल (1) के अंतर्गत फिक्स्ड इफैक्ट परमेटर का अनुमानित और रैनडम इफैक्ट परमेटर का पूर्वकथन है। देखें Rao एवं Molina (2015); Saei एवं Chambers (2003)। पानेलीजेड क्यूसी लिकेलेहूड (PQL) दृष्टिकोण GLMM के आकलन का प्रसिद्ध पद्धति है। यह दृष्टिकोण गैर नॉर्मल चर के लिए लगभग रैखिए फंक्शन का निर्माण करता है और मानता है की यह रैखिए आश्रित चर लगभग नॉर्मल वितरण का पालन करता है। हमलोग इस पद्धति की इस्तेमाल वेरियन्स कॉम्पोनेंट और फिक्स्ड एफफेक्ट परमेटर के आकलन के लिए किए हैं (Breslow एवं Clayton, 1993)। यदपि PQL फिटिंग कभी कभी असंगत और पक्षपाती आकलन देता है परंतु यह पद्धति अनुभव सिद्ध अच्छा कार्य करता है (González-Manteiga et al., 2007)। EPP आकलन परिदृश्य इकाइयों के परीक्षित चर के इकाई मान और अपरिदृश्य इकाइयों के मॉडल (1) आधारित पूर्वानुमानित मानों पर आधारित है। इसके लिए जनसंख्या इकाइयों के सहायक चर की इकाई स्तरीय मान की आवश्यकता होती है। सहायक चर आधारित सूचना की उपलब्ध मात्रा के आधार पर, हमने लघु क्षेत्र अनुपात आकलन के विकल्प अनुभावजन्य पूर्वकथक का प्रस्ताव किया है। इसमें हमने माना है की हमारे पास सहायक चर का मान सिर्फ प्रतिदर्श इकाइयों के लिए उपलब्ध है और अप्रतिदर्श इकाइयों के लिए इसका मान उपलब्ध नहीं है। ऐसी स्थिति प्रायः उन देशों में होता है जहां या तो जनगणना होती ही नहीं है, या यदि होती है तो इसके इकाई स्तर डाटा की प्राप्ति सुलभ नहीं होता है। इस अवस्था में भी मॉडल (1) का अनुकरण कर सकते हैं परंतु अभिव्यक्ति (2) में दिये गए EPP का इस्तेमाल नहीं किया जा सकता। इसके लिए हमने क्षेत्र i के लघु क्षेत्र अनुपात आकलन हेतु एक विशेष एम्पिरिकल प्लग-इन पूर्वकथन (EPP1

द्वारा निर्देशित) का प्रस्ताव करते हैं।

$$\hat{P}_i^{EPP1} = \sum_{j \in S_i} w_{ij} \hat{\mu}_{ij} \quad (3)$$

यहाँ w_{ij} क्षेत्र i में इकाई j के लिए सामान्यीकृत सर्वेक्षण भार है। EPP1 अध्ययनित चर के मान y_{ij} का कृत्रिम मान का इस्तेमाल करता है, जो विभिन्न संबन्धित क्षेत्रों से शक्ति उधार प्राप्त करता है। अतः यह बड़े हुये प्रतिदर्श आकार का उपयोग करता है। आशा करते हैं कि EPP1, DIR के अपेक्षा ज्यादा कुशल होगा। परंतु यह बात पर निर्भर करता है कि कार्यकारी मॉडल वास्तविक मॉडल का कितना अच्छा अनुकरण करता है। EPP1 सहायक चर के मान को प्रतिदर्श भार के बराबर बार दोहराकर सहायक चर के जनसंख्या को प्राप्त करता है। इस आकलन की कमी यह है की यह अध्ययनित चर के वास्तविक मान का इस्तेमाल नहीं करता है। ऐसे क्षेत्र जहां प्रतिदर्श आकार बहुत छोटा नहीं है, वहाँ इसका इस्तेमाल बहुत ज्यादा सुविधाजनक नहीं है। साथ ही साथ EPP1 की गुणवत्ता इस बात पर बहुत ज्यादा निर्भर करता है कि सर्वेक्षण डाटा को मॉडल कितना अच्छा सज्जित करता है। ऐसे क्षेत्र जिसमें प्रतिदर्श आकार शून्य है, उसके लिए EPP1 को लगाना संभव नहीं है।

MSE आकलन

MSE आकलन हेतु हमने अनुमानी विश्लेषण गणितिक सन्निकटन का इस्तेमाल किया है। माना की $x_{ij} = (x_{i1}^T, \dots, x_{iD}^T)^T$ एक $n \times D$ मैट्रिक्स है जिसमें $x_i = (x_{i1}, \dots, x_{iD})^T$ हैं, Z_i एक $n \times D$ डायगोनल मैट्रिक्स है और $u = (u_1, \dots, u_D)^T$ एक नॉर्मल वितरित रैनडम एरिया इफैक्ट का $D \times 1$ वेक्टर है जिसका औसत शून्य और प्रसरण $\Omega = \sigma^2 I_D$ है जिसमें I_D एक $D \times D$ डायगोनल मैट्रिक्स है। अतः लघु क्षेत्र i का अनुपात P_i को इस प्रकार निद्रिस्ट कर सकते हैं $\hat{P}_i^{EPP1} = w_{ij}^T \hat{\mu}_{ij}$; जहां $\hat{\mu}_{ij}, y_{ij}$ के कृत्रिम मानों का n_i वेक्टर है और w_{ij} सामान्यीकृत सर्वेक्षण भार का n_i वेक्टर है जिसका j^{th} अवयव w_{ij} हैं। हम इसे वेक्टर तरीके से भी लिख सकते हैं, $\hat{P}_i^{EPP1} = w_{ij}^T \hat{\mu}_{ij}$, जहां $\hat{P}_i^{EPP1} = (\hat{P}_1^{EPP1}, \dots, \hat{P}_D^{EPP1})^T$, $\hat{\mu}_{ij} = (\hat{\mu}_{i1}, \dots, \hat{\mu}_{iD})^T$ और $w_{ij} = \text{diag}\{w_{ij}; i = 1, \dots, D\}$ सामान्यीकृत सर्वेक्षण



भार का $n \times n$ डायगोनल मैट्रिक्स हैं। माना कि $\hat{\mu}_s = \exp(\hat{\eta}_s) [1 + \exp(\hat{\eta}_s)]^{-1}$ जहाँ $\hat{\eta}_s = x_s \hat{\beta} + z_s u^*$; $\mu_s = \exp(\eta_s) [1 + \exp(\eta_s)]^{-1}$ और $\eta_s = x_s \beta + z_s u$ है। तो, EPP1 के MSE को हम इस प्रकार से निकाल सकते हैं $MSE(\hat{P}^{EPP1}) = w_s E[(\hat{\mu}_s - \mu_s)(\hat{\mu}_s - \mu_s)'] w_s'$ (4)

टेलर सिरीज लिनियरिजेशन द्वारा $\hat{\mu}_s - \mu_s$ का प्रथम स्तर मूल्यांकन है।

$$\hat{\mu}_s - \mu_s = (\hat{\eta}_s - \eta_s)' \left(\frac{\partial \mu_s}{\partial \eta_s} \right)_{\eta_s = \eta_s}$$

इस प्रकार से

$$E[(\hat{\mu}_s - \mu_s)(\hat{\mu}_s - \mu_s)'] = H_s' E[(\hat{\eta}_s - \eta_s)(\hat{\eta}_s - \eta_s)'] H_s$$

जहाँ

$$H_s = \frac{\partial \mu_s}{\partial \eta_s} \Big|_{\eta_s = \eta_s} = \frac{\partial}{\partial \eta_s} \left(\frac{\exp(\eta_s)}{1 + \exp(\eta_s)} \right) \Big|_{\eta_s = \eta_s} = \frac{\exp(\eta_s)}{[1 + \exp(\eta_s)]^2} = \mu_s(1 - \mu_s)$$

है। माना कि $\delta = H_s \hat{\eta}_s = H_s \{x_s \hat{\beta} + z_s u^*\}$ जहाँ $\hat{\delta}$, $\hat{\beta}$ और \hat{u} का रैखिए प्रकलक है, Saei एवं Chambers (2003)।

$$MSE(\hat{P}^{EPP1}) = w_s MSE(\hat{\delta}) w_s' = w_s H_s E[(\hat{\eta}_s - \eta_s)(\hat{\eta}_s - \eta_s)'] H_s' w_s'$$

प्रसाद एवं राव (1990) का परिणाम एवं सन्निकटन का पालन करते हुए EPP1 के MSE का सन्निकटीत अभिव्यक्ति प्राप्त कर सकते हैं। Rao एवं Molina (2015, chapter 5, page 100-107), Saei एवं Chambers (2003), इस प्रकार से

$$MSE(\hat{P}^{EPP1}) = M_1(\sigma^2) + M_2(\sigma^2) + M_3(\sigma^2) \quad (5)$$

यहाँ $M_1(\sigma^2)$ और $M_2(\sigma^2)$ MSE (5) का बड़ा $M_3(\sigma^2)$ भाग है और मानते हैं कि विस्तारण परमेटर σ^2 कि जानकारी प्राप्त है (Rao एवं Molina 2015)। तीसरे भाग $M_3(\sigma^2)$ डाटा द्वारा रैन्डम इफैक्ट के विस्तारण आकलन में परिवर्तनशीलता के कारण है। MSE आकलक के तीनों भाग को इस प्रकार से प्राप्त करते हैं।

माना कि $\Sigma = z_s' L z_s + \Omega^{-1}$ और $L = -\frac{\partial^2 l_s}{\partial \eta_s \partial \eta_s'} \Big|_{\eta_s = \eta_s} = \text{diag} \{ \mu_s(1 - \mu_s) \}$ लॉग

लाइक्लीहूड का $\hat{\eta}_s$ के संदर्भ में द्वितीय स्तर के

डिफरेंसियेशन है। McGilchrist (1994) को मानते हुये

$$V_s = - \begin{bmatrix} x_s' & z_s' \\ z_s' & \Omega^{-1} \end{bmatrix} L \begin{bmatrix} x_s & z_s \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \Omega^{-1} \end{bmatrix} = \begin{bmatrix} x_s' L x_s & x_s' L z_s \\ z_s' L x_s & \Sigma \end{bmatrix} = \begin{bmatrix} v_{s11} & v_{s12} \\ v_{s21} & v_{s22} \end{bmatrix}$$

$$V_s^{-1} = I_{(2k)}^{-1} = \text{Var}(\hat{\beta}, \hat{u})' = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} x_s' L x_s & x_s' L z_s \\ z_s' L x_s & \Sigma \end{bmatrix}^{-1}$$

$\hat{\beta}$ और u के डायमेशन के आधार पर मैट्रिक्स और इसके विपरीत V_s^{-1} को विभाजित कर सकते हैं। यहाँ V_s^{-1} के उप मैट्रिक्स $A_{11} = [x_s' L x_s - x_s' L z_s \Sigma^{-1} z_s' L x_s]$ और $A_{22} = \Sigma^{-1} + \Sigma^{-1} z_s' L x_s A_{11}^{-1} x_s' L z_s \Sigma^{-1}$ है। परिभाषित करें $\Delta = w_s H_s z_s \Sigma^{-1} = z_s' \Sigma^{-1}$ तथा $z_s' = w_s H_s z_s$, माना कि $z_{s(k)}$ मैट्रिक्स z_s' की k वां पंक्ति है इसका डेरिवेटिव

$$\nabla_{(k)} = \frac{\partial \Delta_{(k)}}{\partial \sigma^2} \Big|_{\sigma^2 = \sigma^2} = \frac{\partial (z_{s(k)}' \Sigma^{-1})}{\partial \sigma^2} \Big|_{\sigma^2 = \sigma^2} = (\hat{\sigma}^2)^{-2} z_{s(k)}' \Sigma^{-1} \hat{\sigma}^2$$

हैं। $\Sigma^{-1} = z_s' L z_s + z_s' L z_s \Omega z_s' L z_s$, इस प्रकार से MSE के तीनों भाग को निम्नानुसार से लिख सकते हैं।

$$M_1(\sigma^2) = w_s H_s z_s \Sigma^{-1} z_s' H_s' w_s'$$

$$M_2(\sigma^2) = w_s H_s c_s A_{11} c_s' H_s' w_s' \quad \text{जहाँ}$$

$$c_s = x_s - z_s \Sigma^{-1} z_s' L x_s$$

$$M_3(\sigma^2) = w_s \left[(\nabla_{(k)} \Sigma^{-1} \nabla_{(k)}') \text{var}(\hat{\sigma}^2) \right]$$

यहाँ $\text{var}(\hat{\sigma}^2)$ विस्तारण भाग $\hat{\sigma}^2$ के आकलित मान का अस्थिर विस्तारण हैं। जिसका मान उपयुक्त फिशर सूचना मैट्रिक्स के विपरीत द्वारा प्राप्त की जाती है।

एम्पिरिकल वैल्यूशन

इस भाग में हम विभिन्न आकलकों का तुलनात्मक प्रदर्शन का विवरण प्रस्तुत किया है। विशेष रूप से हमने EPP1 का तुलना DIR जो कि अपनी अस्थिरता के लिए जाना जाता है और EPP जो लघु क्षेत्र आकलन में बहुत ज्यादा उपयोग होता है, से किया गया है। मॉडल आधारित सिमुलेशन हेतु हमने कृत्रिम जनसंख्या निर्माण किया है। हरेक सिमुलेशन के शुरुआत में मॉडल (1) के अंतर्गत जनसंख्या इकाइयों का निर्माण किया गया है तत्पश्चात हरेक स्ट्राटा से एक प्रतिदर्श का चुनाव स्ट्राटिफिएड सिम्पल रैन्डम संप्लिंग द्वारा किया गया है। आकलकों के तुलना हेतु हमने दो मानक रिलेटिव



बायस एवं रिलेटिव रूट मीन स्क्वेर एरर का इस्तेमाल किया है।

$$RB(P) = \text{mean} \left\{ P_i^{-1} \left(K^{-1} \sum_{k=1}^K \hat{p}_{ik} \right) - 1 \right\} \times 100$$

$$RRMSE(P) = \text{mean} \left[P_i^{-1} \left\{ \sqrt{K^{-1} \sum_{k=1}^K (\hat{p}_{ik} - p_{ik})^2} \right\} \right] \times 100$$

इसी प्रकार MSE आकलक के प्रदर्शन के तुलना हेतु निम्न रिलेटिव बायस का इस्तेमाल किया है।

$$RB(M) = \text{mean} \left\{ MSE_i^{-1} K^{-1} \sum_{k=1}^K (mse_{ik} - MSE_i) \right\} \times 100$$

यहाँ k सिम्युलेशन K को निर्देशित करता है तथा P_{ik} लघु क्षेत्र i के k वां सिम्युलेशन के अनुपात को दिखाता है। क्षेत्र i के वास्तविक अनुपात को $P_i = K^{-1} \sum_{k=1}^K p_{ik}$ से तथा वास्तविक MSE को $MSE_i = K^{-1} \sum_{k=1}^K (p_{ik} - P_i)^2$ से निर्देशित करते हैं। True RMSE और Estimated RMSE का सभी क्षेत्रों का औसत मान इस प्रकार से हैं:

$$TRMSE = \text{mean} \left[\sqrt{K^{-1} \sum_{k=1}^K (\hat{p}_{ik} - p_{ik})^2} \right]$$

$$ERMSE = \text{mean} \left[\sqrt{K^{-1} \sum_{k=1}^K mse_{ik}} \right]$$

हमने MSE आकलक का 'two sigma' (*i-nominal 95 percent*) अंतराल आधारित prediction error के normality assumption का अध्ययन किया है, फलस्वरूप Average percent coverage rate को हमने निम्नरूप से प्राप्त किया है।

$$CR(M) = \text{mean} \left[\frac{1}{K} \sum_{k=1}^K I(|\hat{p}_{ik} - p_{ik}| \leq 2\sqrt{mse_{ik}}) \right] \times 100$$

मॉडल बेस्ड सिम्युलेशन हेतु हमने $D=30$ लघु क्षेत्र माना है जिसमें प्रत्येक की जनसंख्या $N_j = 500$ है। यहाँ अध्ययनित चर y_{ij} के मान को GLMM with logit link function द्वारा सृजित किया गया है $y_{ij} = \text{Binomial}(1, \mu_{ij})$ जहाँ $\mu_{ij} = \exp(1 + x_{ij} + u_i) / \{1 + \exp(1 + x_{ij} + u_i)\}^{-1}$; $j = 1, \dots, N_i$; $i = 1, \dots, 30$. यहाँ क्षेत्र विशेष रैंडम इफैक्ट को स्वतंत्र नॉर्मल वितरण औसत 0 और विस्तारण σ^2 से सृजित किया गया है। क्षेत्र विस्तारण के चार मान $\sigma^2 = 0.05, 0.10, 0.20$ और 0.25 का इस्तेमाल

किया गया है। सहायक चर x_{ij} के मानों को स्वतंत्र रूप से one degree of freedom के chi square वितरण से चुना गया है और इसका मान सम्पूर्ण सिम्युलेशन के दौरान स्थिर। हमने simulated population से एक प्रतिदर्श स्ट्राटिफाइड सिम्पल रैंडम संपलिंग द्वारा $n_i = 600$ आकार का चुना, जिसका क्षेत्र प्रतिदर्श आकार $n_{ik} = 20$ स्थिर है। हरेक सिम्युलेशन में जनसंख्या मान स्वतंत्र रूप से सृजित किया गया है तथा हरेक क्षेत्र से प्रतिदर्श स्वतंत्र रूप से चुना गया है। यहाँ $K = 1000$ मॉटे कार्लो सिम्युलेशन किया गया है। हमने सिम्युलेशन को छोटे प्रतिदर्श आकार $n = 300$ एवं 150 पर भी किया है, जिसमें क्षेत्र प्रतिदर्श आकार क्रमशः 10 और 5 है। सिम्युलेशन के लिए हमने R सॉफ्टवेर का इस्तेमाल किया है। जिसमें हमने lme4 library की glmer function को GLMM के फिटिंग के लिए इस्तेमाल किया है। मॉडल बेस्ड सिम्युलेशन आधारित बिन्दु आकलन का परिणाम तालिका 1 में दिया गया है। तालिका 2 में प्रतिदर्श आकार 10 एवं 20 तथा क्षेत्र विशेष विस्तारण प्रभाव $n_{ik} = 0.10$ और 0.20 के औसत MSE मान को प्रदर्शित किया है।

निष्कर्ष

अनुभवजन्य परिणाम दिखाता है कि प्रदित लघु क्षेत्र अनुपात आकलक का प्रदर्शन अन्य आकलकों से अच्छा है एवं इकाई स्तरीय सहायक चर जनसंख्या डाटा के अभाव में यह एक व्यावहारिक विकल्प का कार्य कर सकता है। MSE आकलक का परिणाम भी सराहनीय है। इस संदर्भ में कई ऐसे क्षेत्र हैं जहाँ और अधिक अध्ययन की आवश्यकता है। जैसे, हमने यहाँ माना है कि GLLM में उपयुक्त क्षेत्र विशिष्ट प्रभाव आपस में स्वतंत्र है। परंतु, वास्तविक जीवन में दो पड़ोसी क्षेत्र के इकाइयों का आपस में पूर्णतः अलग होने की संभावना बहुत कम है। अतः प्रदित तरीके का स्थानीय अध्ययन करने की आवश्यकता है।

संदर्भ

ब्रेसलो, एन. ई. एवं क्लेटन, डी. जी. (1993). अप्रोक्सिमेट इन्फेरेंस इन जनरलाइज्ड लिनर मिक्सड मॉडेल्स. जर्नल ऑफ अमेरिकन स्टैटिस्टिक्स



असोशिएशन, 88 (421), 9-25.

चन्द्र, एच., चेंबर्स, आर. एवं सलवाती, एन. (2012). स्माल एरिया एस्टीमेशन ऑफ प्रोपोर्टिओन्स इन बिजनस सर्वे. जर्नल ऑफ स्टैटिस्टिकल कम्प्यूटेशन एण्ड सिम्युलेशन, 82 (6), 783-795.

मैकगिल्वृस्त, सी. ई. (1994). एस्टीमेशन इन जर्नलाइज्ड लिनर मिक्सड मॉडल्स। जर्नल ऑफ रॉयल स्टैटिस्टिकल सोसाइटी, बी 56, 61-69.

प्रसाद, एन.जी.एन. एवं राव, जे.एन.के. (1990). दि एस्टीमेशन ऑफ दि मीन क्युरेड एर इन स्माल एरिया एस्टिमेटोर्स, जर्नल ऑफ अमेरिकन स्टैटिस्टिक्स

एसोशिएशन, 85, 163-171.

राव, जे.एन.के. एवं मोलिना, आई. (2015). स्माल एरिया एस्टीमेशन. जॉन विले एंड, न्यू यॉर्क.

सारण्डल, सी. ई., स्वेंसोन, बी. एवं ब्रेतमन, जे. एच. (1992). मॉडल अरसीस्टेड सर्वे सैम्प्लींग, स्प्रिंगर-वेरलग, न्यू यॉर्क.

सेई, ए. औरचेंबर्स, आर. (2003). स्माल एरिया एस्टीमेशन अंडर लिनर एंड जनरलिज्ड लिनर मिक्सड मॉडल्स विद टाइम एंड एरिया इफेक्ट्स. साउथम्पटन स्टैटिस्टिकल साइन्स रिसर्च इंस्टीट्यूट, एस3आरआई मेथोडॉलोजी वर्किंग पेपर, एम03/15.

Table 1. मॉडल आधारित लघु क्षेत्र अनुपात आकलक EPP1 का प्रतिशत में सभी 30 क्षेत्रों का रेलटिव बायस (RB) और रेलटिव रूट मीन स्क्वेर एर (RRMSE)

σ^2	Predictors	$n_i = 5$		$n_i = 10$		$n_i = 20$	
		RB (P)	RRMSE(P)	RB (P)	RMSE(P)	RB (P)	RRMSE(P)
0.05	DIR	-0.11	19.94	0.15	14.06	0.01	9.81
	EPP	0.37	6.13	0.38	5.05	0.14	4.30
	EPP1	0.31	7.97	0.39	6.15	0.15	4.94
0.1	DIR	-0.07	20.00	0.15	14.11	0.01	9.86
	EPP	0.46	7.03	0.44	6.00	0.20	5.21
	EPP1	0.39	8.71	0.44	6.98	0.21	5.78
0.2	DIR	-0.10	20.22	0.15	14.21	0.02	9.89
	EPP	0.54	8.51	0.55	7.40	0.29	6.34
	EPP1	0.47	10.02	0.55	8.26	0.30	6.85
0.25	DIR	-0.11	20.30	0.12	14.25	0.02	9.91
	EPP	0.59	9.12	0.57	7.92	0.32	6.72
	EPP1	0.53	10.55	0.57	8.75	0.34	7.21

Table 2- EPP1 आकलक के मॉडल आधारित डैम्के सभी क्षेत्रों का प्रतिशत में औसत RB, RRMSE और coverage rates (CR).

n_i	σ^2	RB(M)	RRMSE(M)	CR(P)
10	0.10	6.17	13.00	94
	0.20	-1.73	5.80	92
20	0.10	-4.05	8.95	92
	0.20	-6.22	11.97	92

सर्वेक्षण डाटा विश्लेषण में सर्वेक्षण भार का महत्व

वन्दिता कुमारी, हुकुम चन्द्र, सुशील कुमार, प्रदीप बसाक, अंकुर विश्वास एवं कौत्सव आदित्य

सार

यदि नमूना का चयन सामान्य यादृच्छिक प्रतिदर्श सर्वेक्षण के बजाय जटिल प्रतिचयन डीजाइन द्वारा किया गया हो तो प्रतिचयनित इकाइयों की चयन होने की प्रायिकताएं भिन्न होती हैं। अतः साधारण यादृच्छिक सर्वेक्षण डाटा जैसा किया गया आकलन समिष्ट मानदंडों का पक्षपातपूर्ण अनुमान प्रदान करता है। इस लेख में हमने सर्वेक्षण भार की उपयोगिता एवं प्रभाव की विभिन्न पहलुओं का विस्तृत विवरण किया है।

परिचय

प्रतिदर्श सर्वेक्षणों में अधिकतर जटिल प्रतिचयन डिजाइनों का उपयोग किया जाता है जिसमें प्रतिचयन इकाइयों की चयनित होने की प्रायिकताएं भिन्न होती हैं। ये जटिल प्रतिदर्श सामान्य यादृच्छिक प्रतिदर्श सर्वेक्षण (simple random sampling) पद्धति द्वारा चयनित प्रतिदर्शों से अलग होते हैं। सामान्य यादृच्छिक प्रतिदर्श के मामले में प्रत्येक इकाइयों को प्रतिदर्श में चयनित होने की प्रायिकता समान होती है और एक इकाई का चयन दूसरे इकाई के चयन पर निर्भर नहीं करता है। यदि प्रतिदर्श का चयन समूहबद्ध या स्तरीकरण जैसे जटिल डिजाइन द्वारा किया गया हो तो इकाइयों एक दूसरे से स्वतंत्र नहीं होती हैं। जिस कारण प्रायः दो मुद्दे उत्पन्न होते हैं, प्रथम, प्रतिदर्श का प्रस्तुतीकरण और मानदंडों (पैरामीटर) के आकलन पर इसके प्रभाव या दूसरा, समिष्ट के विचरण और मानक त्रुटियों का आकलन। यदि जटिल सर्वेक्षण द्वारा एकत्र किए गए डाटा का विश्लेषण साधारण यादृच्छिक सर्वेक्षण द्वारा एकत्रित डाटा जैसा किया जाता है तो यह आकलन समिष्ट मानदंडों का पक्षपातपूर्ण अनुमान प्रदान करता है और मानक त्रुटि को कम करके अनुमानित करता है। अतः ऐसे मामलों में सर्वेक्षण भार का इस्तेमाल लक्ष्य समिष्ट के लिए प्रतिनिधि प्रतिदर्श

बनाने एवं विचरण कम करने के लिए किया जाता है। चयनित इकाई का सर्वेक्षण भार w_i उपयुक्त प्रतिचयन डिजाइन के आधार पर चयनित होने की प्रायिकता का व्युत्क्रम होता है। सर्वेक्षण भार की व्याख्या प्रतिदर्श इकाई द्वारा दर्शाए गए जनसंख्या इकाइयों की संख्या के रूप में किया जाता है। उदाहरण के तौर पर, यदि किसी इकाई के लिए सर्वेक्षण भार 100 इंगित है तो वह इकाई खुद को एवं (99 अन्य इकाइयों) लक्षित समिष्ट के 99 अन्य इकाइयों का प्रतिनिधित्व करता है। कुछ विशेष मामलों को छोड़कर, सभी इकायों का भार $w_i \geq 1$ होता है, क्योंकि चयनित इकाई कम से कम खुद का प्रतिनिधित्व जरूर करता है। सभी गैर-नमूना इकाइयों के लिए $w_i = 0$ होता है। यदि जाने अनजाने डाटा विश्लेषक यह मानता है कि नमूना बिना किसी गैर-उत्तरदायित्व या कवरेज त्रुटि के साथ साधारण यादृच्छिक विधि द्वारा चयनित किया गया है और सर्वेक्षण भार की उपेक्षा करता है यानि सर्वेक्षण भार $w_i = 1$ मानता है तो ऐसा विश्लेषण एक पक्षपातपूर्ण अनुमानक प्रदान करता है। क्योंकि ऐसे विश्लेषण में नमूनाकरण के असमान प्रायिकता प्रभाव को अनुमानक में समाहित नहीं किया जाता है। उदाहरण के लिए, यदि प्रतिदर्श का आकार n है और यदि सर्वेक्षण भार पर ध्यान नहीं जाय, तो समिष्ट के मध्यमान का आकलन निम्न प्रकार से किया जाता है।

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

जहां पर y_i , i^{th} प्रतिदर्श इकाइयों का प्रेक्षण मान है। \bar{y} के सर्वेक्षण भार सहित अनुमानक को निम्न प्रकार से प्रदर्शित किया जाता है।

$$\bar{y}_w = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i}$$

यह दर्शाया जा सकता है कि \bar{y}_w एक निष्पक्ष



अनुमानक है, जबकि \bar{y} एक पक्षपातपूर्ण अनुमानाक है जिसे $\frac{-cov(Y, W)}{W}$ से दर्शाया जा सकता है, जहाँ W भार का मध्यमान है (किश, 1987, 1992)। चूंकि यहाँ सर्वेक्षण भार प्रेक्षण के साथ सहसंबंधित हैं, अतः यह एक पक्षपातपूर्ण अनुमानक है।

भार की अवधारणा

सर्वेक्षण भार के विभिन्न घटकों की प्रेरणा को समझने के लिए असमान प्रायिकता प्रतिचयन और सहायक सूचना के उपयोग के माध्यम से अनुमान सटीकता में सुधार जैसे कारणों को समझने की आवश्यकता है।

असमान प्रायिकता प्रतिचयन

उदाहरण के लिए, ऐसे क्षेत्र के सर्वेक्षण पर विचार करते हैं जहाँ 10% निवासी ग्रामीण समुदाय से हैं और 90% शहरी समुदाय से हैं। समान प्रायिकता प्रतिचयन पद्धति के अनुसार अगर 5000 आकार के निवासी प्रतिदर्श का चयन करना है तो ग्रामीण समुदाय से लगभग 500 निवासी और शहरी समुदाय से 4500 निवासी का चयन करना होगा। ग्रामीण और शहरी दोनों क्षेत्र की विशेषता अनुमानाकों के लगभग समान परिशुद्धता प्राप्त करने हेतु पूर्ण आबादी को ग्रामीण और शहरी क्षेत्रों में विभाजित कर, यदि 2500 निवासियों को प्रत्येक स्तर से चुना जाय तो इससे लक्ष्य आबादी के अनुमान में समस्याएं उत्पन्न हो जाती हैं। क्योंकि इसमें ग्रामीण निवासियों की संख्या का अति-प्रतिनिधित्व होगा है जो की नियमित समान प्रायिकता प्रतिदर्श का पांच गुना है और शहरी निवासियों का प्रतिनिधित्व नियमित प्रतिनिधित्व का 5/9 भाग ही होगा। अतः इस प्रकार से कुल आबादी हेतु निकाला गया निष्कर्ष, ग्रामीण समुदाय की ओर पक्षपातपूर्ण है। इस त्रुटि से निपटने हेतु, सर्वेक्षण भार का उपयोग किया जाता है जिससे यह सुनिश्चित होता है कि ग्रामीण समुदाय आबादी के अनुमान में 10% योगदान प्रदान करे और शहरी समुदाय 90% योगदान करे।

प्रतिचयन विचरण में कमी

पोस्ट-सर्वेक्षण भार अक्सर ऐसे कारकों को शामिल

करता है जो अनुपात आकलन और पोस्ट-स्तरीकरण के माध्यम से विचरण को कम करता है। सहायक चर का भार निर्माण में उपयोग अनुमानक की दक्षता को बढ़ावा देता है। इसके लिए अध्यनरत चर, Y , और सहायक चर, X , का सहसंबंधित होना आवश्यक है। उदाहरणार्थ आबादी की औसत आय का अनुमान हेतु यदि हमें पांच साल पहले का प्रत्येक सदस्य की आय जानकारी हो तो इस जानकारी को सहायक डाटा के गठन हेतु इस्तेमाल कर सकते हैं। यदि सामान्य यादृच्छिक प्रतिदर्श सर्वेक्षण द्वारा प्रत्येक व्यक्ति की वर्तमान आय Y प्राप्त हुआ है। \bar{y} सहायक डाटा का मध्यमान है; \bar{y} मौजूदा सर्वेक्षण डाटा का मध्यमान है और \bar{x} सहायक आंकड़ों के आधार पर यदि जनसंख्या का औसत आय दर्शाता है तो इस स्थिति में अनुमानक को निम्न प्रकार से परिभाषित करते हैं।

$$\bar{y}_e = \frac{\bar{y}}{\bar{x}} \bar{X} = \sum_{i=1}^n w_{hi} y_i$$

जहाँ, $w_{hi} = \frac{1}{n} \frac{X_i}{\bar{x}}$ एक अनुपातिक भार है। क्योंकि पिछले आय डाटा, X , और वर्तमान आय डाटा, Y , का सहसंबद्ध होने की बहुत अधिक संभावना है, अतः \bar{y}_e , \bar{y} की तुलना में \bar{y} के अधिक सटीक अनुमानक है (लोहर, 1999)।

पोस्ट-स्तरीकरण अनुपात आकलन और स्तरीकरण दोनों की निकटता से संबंधित है। जैसा कि नाम से ही निद्रिस्ट है, इसमें इकाइयों का स्तरीकरण विशेषता के आधार पर किया जाता है जिसका ज्ञान हमें डाटा संग्रह के बाद होता है। अनुपात आकलन के समान ही पोस्ट-स्तरीकरण के लिए भी आवश्यक है कि पोस्ट-स्तरीय कुल का ज्ञान निश्चित रूप से हो।

माना कि आबादी में कुल तीन प्रकार के घर आकार है और प्रत्येक प्रकार के घरों की संख्या ज्ञात है, जिसे N_h ; $h=1,2,3$ द्वारा निर्दिष्ट किया गया है। प्रत्येक घर के औसत आय \bar{y}_h और दी n_h प्रतिदर्श में प्रत्येक आकार के घरों की संख्या को n_h निर्देशित करता हो तो इसका पोस्ट-स्तरीकरण आय अनुमानक निम्न होता है।

$$\bar{y}_{ps} = \sum_{h=1}^3 \frac{N_h}{N} \bar{y}_h = \sum_{i=1}^n w_{(h,i)} y_i$$



जहाँ, $w_{mi} = \frac{N_i}{N} \frac{1}{n_i}$ चरण 'h' के यूनिट i के पोस्ट स्तरीकरण भार है।

सर्वेक्षण के भार का निर्माण

मान लें कि n व्यक्तियों का एक प्रतिदर्श के चयन हेतु प्रायिकता Π_i , $i = 1, \dots, n$ है तो अधिकांश सर्वेक्षण स्थितियों में, सर्वेक्षण भार w_i , तीन घटकों का बने होते हैं, इसमें पहला घटक, w_{Bi} , सर्वेक्षण डिजाइन से अर्जित डिजाइन भार है और यह इकाई के चयन प्रायिकता का प्रतिलोम है।

अन्य दो भार डिजाइन भार के सुधार के लिए हैं जिनमें प्रथम, गैर-प्रतिक्रिया के क्षतिपूर्ति करने के लिए और दूसरा, नॉन कवरेज के क्षतिपूर्ति करने के लिए है। दूसरे भार को पोस्ट स्ट्रैटीफिकेशन एडजस्टमेंट फैक्टर भी कहा जाता है, इस भार में प्रतिदर्श विचरण को कम करने हेतु अनुपात समायोजन कारकों को भी शामिल किया जाता है।

डिजाइन भार

डिजाइन भार के लिए प्रतिदर्श इकाइयों के चयन प्रायिकता का प्रतिलोम लिया जाता है (होर्विज एवं थोम्सन, 1952) $w_{mi} = \frac{1}{\pi_i}$ जहाँ, π_i i^{th} प्रतिदर्श की चयन प्रायिकता है। समान प्रायिकता वाले चयन विधि में प्रतिदर्श के प्रत्येक इकाई का सर्वेक्षण भार $\pi_i = \frac{N}{n}$ होता है। स्तरीकृत प्रतिचयन विधि में प्रतिदर्श प्रत्येक स्तर में समान प्रायिकता से चयनित होता है, इसमें स्ट्रेटम h में i^{th} इकाई की चयन प्रायिकता n_{hi}/N_h है जहाँ N_h स्ट्रेटम आकार है और n_{hi} स्ट्रेटम h का प्रतिदर्श आकार है। इस प्रकार से, स्ट्रेटम h में i^{th} इकाई का डिजाइन भार इस प्रकार है।

$$w_{Bhi} = \frac{N_h}{n_{hi}}$$

बहुचरणीय प्रतिचयन विधि में चयन प्रायिकता का अभिकलन प्रत्येक चरण के चयन प्रायिकताओं के गुणनफल के रूप में किया जाता है। तीन चरणीय प्रतिचयन में, प्रथम चरण i ; द्वितीय चरण j के इकाई k की चयन प्रायिकता निम्न प्रकार से दर्शाया जाता है, $P(i, j, k)$ इकाइयों का चयनित होना)

$= P$ (प्रथम चरण में i का चयन होना) $\times P$ (द्वितीय चरण से j का चयन होना प्रथम चरण में i चयनित है) $\times P$ (तृतीय k चरण में k का चयन होना द्वितीय चरण का j प्रथम चरण के i से चयनित है)

यदि प्रत्येक चरण में इकाइयों प्रतिस्थापन के साथ समान प्रायिकता से चयन हुआ हो तो

$$P(\text{unit } i, j, k \text{ selected}) = \frac{n_i n_j n_k}{N N_i N_j}$$

जहाँ प्रथम चरण के N इकाइयों से n इकाइयों का चयन हुआ है, उसके बाद प्रत्येक चयनित प्रथम चरण इकाई i के N_i इकाइयों से द्वितीय चरण का n_i इकाई चुना गया है और प्रत्येक j चयनित तृतीय चरण के N_{ij} इकाइयों में से n_{ij} इकाइयों का चयन किया गया है। तदनुपरान्त इसका भार निम्न प्रकार से होता है।

$$w_{Bijk} = \frac{N N_i N_j}{n_i n_j n_k}$$

स्तरीकृत दो चरण डिजाइन के लिए अगर प्राथमिक प्रतिदर्श इकाइयों जिले हो और दूसरी चरण की इकाइयों उन जिलों के ऐसे बच्चे हो जो सर्वेक्षण पात्रता मानदंडों को संतुष्ट करते हैं तो i^{th} जिले के j^{th} बच्चे का प्रतिदर्श में चयन की प्रायिकता, जिला के चयन की प्रायिकता π_{hi} एवं चयनित जिले के बच्चे की चयन प्रायिकता का गुणनफल होता है और इसका प्रतिलोम सर्वेक्षण भार होता है।

$$w_{Bhidj} = \frac{1}{\pi_{hi} \pi_{hij}}$$

एक स्तर के भीतर सभी प्रतिदर्श इकाइयों के लिए वजन का योग उस स्तर के भीतर योग्य बच्चों की कुल संख्या का अनुमान देता है और इन स्तरों का योग जनसंख्या में सभी योग्य बच्चों का अनुमान देता है। अगर प्रतिस्थापन के बिना आकार की आनुपातिक प्रायिकता के द्वारा जिलों का चयन किया गया हो तो जिलों के लिए चयन प्रायिकता $n_{hi} N_{hi} / N_i$ होता है, जहाँ n_{hi} स्ट्रेटम h से चयनित प्रथम चरण इकाइयों की संख्या है, N_{hi} जिला में बच्चों की संख्या है और N_i स्तर h में बच्चों की कुल संख्या है। बच्चे (h, i, j) की चयन प्रायिकता को n_{hij} / N_{hi} के रूप में गणना की जाती है, जहाँ n_{hij} i^{th}



जिलों में चयनित बच्चों की संख्या है और N_{hi} जिले में बच्चों की संख्या है। इन परिभाषाओं के अंतर्गत, इसका सर्वेक्षण भार निम्न है।

$$w_{hij} = \frac{N_h}{n_h N_{hi}} \times \frac{N_{hi}}{n_{hi}}$$

पोस्ट-स्तरीकरण भार

डिजाइन भार द्वारा समावेश प्रायिकताओं में अंतर को वर्णित कर डिजाइन पक्षपात को सही करता है। पोस्ट-स्तरीकरण भारोत्तोलन ऐसी विधि है जो सहायक सूचना के उपयोग द्वारा प्रतिचयानात्मक त्रुटि और संभावित गैर-प्रतिसाद पूर्वाग्रह को कम करता है। यहाँ डिजाइन भार को समायोजित कर पोस्ट-स्तरीकरण भार को इस तरह बनाते हैं कि यह जनसंख्या में आयु वर्ग, लिंग और शिक्षा के क्रॉस वर्गीकरण का वितरण एवं जनसंख्या में क्षेत्र के सीमांत वितरण को दोहराता हो। पोस्ट स्तरीकरण भार का डिजाइन भार पर निम्न लाभ है: प्रथम, अगर अध्ययनित चर और पोस्ट-स्ट्रैटिफिकेशन चर के बीच रैखिक निर्भरता हो तो यह प्रतिचयानात्मक त्रुटि को कम करता है। दूसरा, यदि अध्ययनित चर पर प्रतिक्रिया देने की प्रवृत्ति एवं पोस्ट-स्तरीकरण चर के बीच रैखिक निर्भरता हो तो यह मौजूद गैर प्रतिक्रिया पूर्वाग्रह को कम करता है।

गैर-प्रतिक्रिया का समायोजन

किस भी सर्वेक्षण में यह दुर्लभ है कि सभी सर्वेक्षण इकाइयों से सारी वांछित जानकारी प्राप्त हो सके। प्रायः ऐसा होता है कि कुछ घरों में कोई भी डाटा उपलब्ध ही नहीं हो (यूनिट या कुल गैर-प्रतिक्रिया), या केवल आंशिक डाटा उपलब्ध हो, अर्थात्, सर्वेक्षण में कुछ सवालों के उत्तर ही प्रदान नहीं किया गया हो (इकाई गैर प्रतिक्रिया)। इस तरह के सर्वेक्षणों पर आधारित आकलन प्रायः पक्षपातपूर्ण होता है। इसलिए, सर्वेक्षण डाटा में गैर-प्रतिक्रिया को निम्न स्तर पर रखना बहुत ही महत्वपूर्ण है। उदाहरण के लिए, शहरी क्षेत्रों में अपेक्षाकृत अधिक आय वाले लोगों की किसी बहुउद्देशीय सर्वेक्षण में शामिल होने की संभावना कम होती है। इस तरह की गैर-भागीदारी औसत घरेलू

आय, शैक्षिक प्राप्ति, साक्षरता आदि के राष्ट्रीय अनुमान को भी प्रभावित करता है।

दुनिया भर के प्रमुख घरेलू सर्वेक्षणों में गैर-प्रतिक्रिया हेतु सर्वेक्षण भार का समायोजन एक महत्वपूर्ण अभ्यास है। यहाँ गैर-उत्तरदायी प्रतिदर्श इकाइयों के आधार पर समायोजन भार का स्थानांतरण उत्तरदायी इकाइयों को निम्नलिखित चरणों में किया जाता है:

- I. प्रारंभिक भार लागू करना
- II. नमूना को उपसमूहों में विभाजित कर प्रत्येक उपसमूह के लिए भारित प्रतिक्रिया दरों की गणना करना
- III. गैर-प्रतिक्रिया समायोजन के लिए उप-समूह प्रतिक्रिया दर के पारस्परिक का उपयोग करना
- IV. i^{th} इकाई के लिए गैर-प्रतिक्रिया समायोजित वजन की गणना करना:

$$w_i = w_{i1} * w_{i2}$$

जहाँ w_{i1} प्रारंभिक भार एवं w_{i2} गैर-प्रतिक्रिया समायोजन भार है। भारित गैर-प्रतिक्रिया दर को योग्य चयनित मामलों में पूर्ण साक्षात्कार की भारित संख्या और चयनित मामलों की भारित संख्या का अनुपात के रूप में परिभाषित करते हैं।

पोस्ट-स्तरीकरण समायोजन

गैर-कवरेज लक्षित समिष्ट को कवर हेतु उपयुक्त प्रतिचयन फ्रैम की विफलता को दर्शाता है और इस प्रकार कुछ इकाइयों का घरेलू सर्वेक्षण हेतु चयनित प्रतिदर्श में चयन होने की प्रायिकता शून्य होती है। गैर-कवरेज प्रायः दुर्गम क्षेत्रों की वजह से होती है साथ ही, दूरदराज के परिवारों को कभी-कभी घरेलू सर्वेक्षणों के लिए प्रति चयन फ्रैम से निकाल दिया जाता है क्योंकि उनको कवर करना काफी महंगा होता है। इसके अलावे भी कई अन्य कारणों से गैर-कवरेज हो सकता है। घरेलू सर्वेक्षणों में होने वाले गैर कवरेज की समस्या को निपटाने हेतु कई प्रक्रियाएं हैं। जैसे कि, कई फ्रैम, बेहतर सूचीबद्ध प्रक्रियाओं के उपयोग एवं भार के सांख्यिकीय समायोजन के माध्यम से गैर-कवरेज हेतु क्षतिपूर्ति को शामिल है। यदि संपूर्ण



समिष्ट एवं जनसंख्या के निर्दिष्ट उपसमूहों हेतु विश्वसनीय नियंत्रण योग उपलब्ध हो, तो ऐसे प्रतिचयन इकाइयों के भार को समायोजित करने का प्रयास किया जा सकता है, ताकि निर्दिष्ट उपसमूहों में नियंत्रण योग के भार का योग बनाया जा सके। यहाँ उपसमूह को पोस्ट-स्ट्रेटम तथा सांख्यिकीय समायोजन प्रक्रिया को पोस्ट स्तरीकरण कहा जाता है। यह प्रक्रिया एक साथ गैर-प्रतिक्रिया और गैर-कवरेज द्वारा क्षतिपूर्ति को काबू करता है। यह एक दिये हुए चर हेतु भारित प्रतिचयन वितरण को समायोजित करता है ताकि ज्ञात समिष्ट वास्तविक वितरण के अनुरूप हो।

विचरण पर भार का प्रभाव

कुछ जनसंख्या उपसमूहों की ओवर सेमिप्लिंग और पोस्ट-सर्वे समायोजन के कारण भार प्रतिदर्शों में व्यापक भिन्नता होती है, जो अक्सर अनुमानों की परिशुद्धता को प्रभावित करता है। परिशुद्धता पर असमान भार के गहन प्रभाव के कारण, इसका ध्यान सर्वेक्षण डिजाइन चरण में ही रखना चाहिए। असमान भार के कारण विचरण स्फीति को समायोजित करने के लिए प्रतिचयन आकार को बढ़ाया जा सकता है। यदि भार की भिन्नता बहुत चरम है, तो प्रतिचयन के लिए योजना ध्यान पूर्वक बनाने की आवश्यकता है।

समश्रायण विश्लेषण में भार का उपयोग

माना कि समिष्ट U सुपरपॉपुलेशन से सृजित एक N आकार का यादृच्छिक प्रतिदर्श है

$$y_k = \alpha + \beta x_k + e_k$$

जहाँ $e_1, \dots, e_k \sim (0, \sigma^2)$ परिमित समिष्ट के पैरामीटर β का सुपर पॉपुलेशन मॉडल पैरामीटर के अनुमानक के रूप में देखा जा सकता है (सर्नडल इत्यादि, 1992)

$$\beta = \frac{\sum_{k \in U} (y_k - \bar{Y})(x_k - \bar{X})}{\sum_{k \in U} (x_k - \bar{X})^2}$$

प्रतिगमन गुणांक के सुज्ञात सामान्य नुन्यतम वर्ग (OLS) अनुमानक को निम्न द्वारा दिया जाता है

$$\hat{\beta} = \frac{\sum_{k \in s} (y_k - \bar{y})(x_k - \bar{x})}{\sum_{k \in s} (x_k - \bar{x})^2}$$

जहाँ $x = \frac{\sum_{k \in s} x_k}{n}$ और $y = \frac{\sum_{k \in s} y_k}{n}$ है। प्रतिगमन गुणांक के भारित सुज्ञात सामान्य नुन्यतम वर्ग (WOLS) अनुमानक को निम्न द्वारा दिया जाता है

$$\hat{\beta} = \frac{\sum_{k \in s} w_k (y_k - \bar{y})(x_k - \bar{x})}{\sum_{k \in s} w_k (x_k - \bar{x})^2}$$

जहाँ $\bar{x} = \frac{\sum_{k \in s} w_k x_k}{\sum_{k \in s} w_k}$ और $\bar{y} = \frac{\sum_{k \in s} w_k y_k}{\sum_{k \in s} w_k}$ है।

समाश्रयण विश्लेषण में भारित डाटा का उपयोग की दो स्थितियाँ हो सकती हैं: एक, जहाँ मॉडल में शामिल समश्रायक चर भार का फंक्शन है और दूसरा जहाँ भार आश्रित चर अर्थात् त्रुटियों का फंक्शन है। पहली स्थिति में, प्रतिचयन भार WOLS और OLS निष्पक्ष और सुसंगत पैरामीटर का अनुमान प्रदान करते हैं परन्तु, OLS छोटे मानक त्रुटि के साथ अनुमान प्रदान करता है। इसलिए, यहाँ OLS का उपयोग किया जायेगा और दूसरे मामले में, WOLS का उपयोग कर मॉडल का अनुमान उचित होगा।

निष्कर्ष

पहले प्रकाशित अध्ययनों के बावजूद, सर्वेक्षण डाटा के विश्लेषण हेतु भार का उपयोग हमेशा नहीं किया जाता है। भले ही डाटा जटिल सर्वेक्षण डिजाइन से क्यों न हो, सांख्यिकी विद कई कारणों से जैसे सरलता, बिना सर्वेक्षण भार के डाटा का विश्लेषण करते हैं। जटिल सर्वेक्षण में डाटा विश्लेषण हेतु सर्वेक्षण भार का उपयोग नहीं करना एक पक्षपाती और गलत निष्कर्ष प्रदान करता है। पोस्ट स्ट्रैटिकेशन वजन को छोड़कर, प्रतिचयन भार विचरण को बढ़ाते हैं। जटिल नमूने आमतौर पर एक ही आकार की समान समिष्ट के सामान्य यादृच्छिक प्रतिदर्श सर्वेक्षण की तुलना में बड़े मानक त्रुटि उत्पन्न करते हैं। गैर प्रतिक्रिया पूर्वाग्रह को कम करने हेतु प्रतिचयन भार का भी इस्तेमाल किया जा सकता है। वर्तमान में प्रतिचयन भार को सर्वेक्षण आंकड़ों के विश्लेषण का एक अभिन्न अंग माना जाता है। अधिकांश सर्वेक्षण कार्यक्रम अब भार के उपयोग



की वकालत करता हैं। अतीत में कंप्यूटिंग इंफ्रास्ट्रक्चर के कमी के कारण भी सर्वेक्षण डिजाइनर स्वभार वाले प्रतिदर्शों को ज्यादा उपयोग करते थे तथा सर्वेक्षण डाटा के विश्लेषण में भार का इस्तमाल नहीं किया करते थे। हालांकि, कंप्यूटर प्रौद्योगिकी का विकास एवं विशेषीकृत कंप्यूटर सॉफ्टवेयर पैकेज के सृजन के कारण सर्वेक्षण डाटा के विश्लेषण काफी उपयोगी सिद्ध हुए हैं।

सन्दर्भ:

किश. (1987): स्टैटिस्टिकल डिजाइन फॉर रिसर्च, न्यू यॉर्क: विले

किश. (1992): "वेइंग फॉर उनइकवल P_1 ", जर्नल ऑफ ऑफिशियल स्टैटिस्टिक्स, 8, 183-200

सर्नडल सी.ई., स्वेनस्सन, बी. एवं रिटमैन, जे. (1992): "माडल असिस्टेड सर्वे सैंपलिंग", रिप्रिन्गेर वरलेग, न्यू यॉर्क

लोहर. (1999): "सैंपलिंग डिजाइन एंड एनालिसिस", पसिफिक ग्रोव, सी.ए.: दुक्सबुरी प्रेस

होर्वित्ज, डी.जि. एवं थॉम्सन, डी.जे. (1952): "ए जेनरलाईसेसंन ऑफ सैंपलिंग विथौट रिप्लेसमेंट फ्रॉम ए फाईनाइट यूनिवर्स", जर्नल ऑफ दी अमेरिकन स्टैटिस्टिकल एसोसिएशन, 47, 663-685

रैपिड माइजर के द्वारा टेक्स्ट माइनिंग

कृष्ण कुमार चतुर्वेदी, अनिल राय, मो. समीर फारूकी, शशि भूषण लाल,

अनु शर्मा, यू. बी. अंगडी एवं विशाल मिश्र

प्रस्तावना

टेक्स्ट माइनिंग टेक्स्ट या टेक्स्ट संबंधी स्रोतों के दिए गए कॉरपोरा से जानकारी और ज्ञान निकालने के बारे में है। टेक्स्ट माइनिंग उपलब्ध ग्रंथों से अज्ञात ज्ञान को खोजने के लिए दस्तावेजों के बड़े सेट को संकलित, व्यवस्थित और विश्लेषित करने की प्रक्रिया है, जो विश्लेषकों और निर्णय निर्माताओं के लिए उपयोगी होती है। सरल शब्दों में, यह उपलब्ध टेक्स्ट से अनदेखे ज्ञान की खोज के रूप में परिभाषित भी किया जा सकता है।

टेक्स्ट माइनिंग डेटा माइनिंग के आगामी महत्वपूर्ण क्षेत्रों में से एक है जिसका उद्देश्य उपलब्ध टेक्स्ट डेटा या दस्तावेजों से उपयोगी और नई जानकारी ढूँढना या खोजना है।

डेटा माइनिंग डेटा सेटों, चरों का चयन, डेटा तैयार करने और नवाचार जानकारी या ज्ञान को खोजने के लिए प्रवृत्तियों और वितरण का विश्लेषण करने के रूप में परिभाषित किया गया है। इसके अतिरिक्त, टेक्स्ट माइनिंग दस्तावेजों या दस्तावेजों के सेट की पहचान करने, चरों को निकालने, उपयुक्त डेटा कटौती तकनीकों का उपयोग करते हुए, डेटा सेट तैयार करने के बाद उपयोगी या महत्वपूर्ण विशेषताओं का चयन करने और फिर छुपी हुई जानकारी / ज्ञान को उजागर करने के लिए वाक्यात्मक और अर्थपूर्ण तरीके का उपयोग करते हुए इसमें उपलब्ध रुझानों का विश्लेषण करने के लिए भी संदर्भित करता है।

टेक्स्ट माइनिंग पूरी तरह से मूल भाषा पर निर्भर है और नेचुरल लैंग्वेज प्रोसेसिंग तकनीकों का उपयोग करके प्राप्त किया जा सकता है। टेक्स्ट डेटा में शब्द, वाक्य, और पैराग्राफ का समूह शामिल होता है

जिसमें इसका उद्देश्य और उपयोग होता है। किसी भी विश्लेषणात्मक तकनीक में यह वांछनीय है कि टेक्स्ट डेटा को विश्लेषित करने के लिए उसे संख्यात्मक रूप में परिवर्तित करने की आवश्यकता होती है। टेक्स्ट डेटा का अनुप्रयोग विभिन्न क्षेत्रों जैसे फीडबैक प्रोसेसिंग, ओपिनियन माइनिंग, टेक्स्ट सार (समरी), समाचार रिपोर्ट विश्लेषण, स्पैम फिल्टरिंग और पहचान आदि समानता, पुनःप्राप्ति, और अन्य उपलब्ध ज्ञान की खोज से शुरू होता है। दस्तावेज में आमतौर पर कई प्रासंगिक और अप्रासंगिक जानकारी होती है। अप्रासंगिक या अनुपयोगी जानकारी को दस्तावेजों या रिपोर्ट के पूर्व प्रसंस्करण की विभिन्न तकनीकों के द्वारा हटाया जा सकता है। टेक्स्ट माइनिंग के दो चरणों को दस्तावेजों के पूर्व प्रसंस्करण और माइनिंग तकनीक के अनुप्रयोग में विभाजित किया गया है।

ओमिक्स जैविक विज्ञान, प्रोटीओमिक्स या मेटाबोलमिक्स जैसी जीव विज्ञान के अंतर्गत आने वाले क्षेत्रों को संदर्भित करता है। ओमिक्स एक सामान्य शब्द है जो की जीव विज्ञान के विभिन्न विषय जैसे जीनोम, प्रोटीम या मेटाबोलोम में अध्ययन करने के लिए उपयोग किया जाता है। ओमिक्स डेटा में टेक्स्ट माइनिंग को बायोमैडिकल टेक्स्ट माइनिंग (जिसे बायोएनएलपी भी कहा जाता है) के रूप में संदर्भित किया गया है। बायोएनएलपी में, टेक्स्ट माइनिंग बायोमैडिकल और आणविक जीव विज्ञान ग्रंथों और उनसे सम्बंधित साहित्य पर लागू होता है। यह नेचुरल लैंग्वेज प्रोसेसिंग, जैव सूचना विज्ञान, चिकित्सा सूचना विज्ञान और कम्प्यूटेशनल भाषाविज्ञान का एक नया शोध क्षेत्र है।



टेक्स्ट डेटा का पूर्व प्रसंस्करण

किसी भी टेक्स्ट दस्तावेज के पूर्व प्रसंस्करण के लिए मानक निम्न चरणों में शामिल हैं :

टोकनकरण : टोकनकरण शब्दों एवं अक्षरों की एक धारा को टोकन के क्रम में परिवर्तित करने की प्रक्रिया को दर्शाता है। टोकन एक वाक्य, पैराग्राफ, एक पंक्ति, एक शब्द या वर्णमाला हो सकती है। टोकनकरण का उद्देश्य दिए गए टेक्स्ट को अलग-अलग करना है। यहां, हम टोकन को एक शब्द या एक वाक्य के रूप में देखते हैं। हम सभी विराम चिह्न, गैर-प्रिंट करने योग्य वर्ण और अन्य अर्थहीन प्रतीकों को हटा देते हैं क्योंकि वे वर्गीकरण कार्य में सामान्यतः योगदान नहीं करते हैं। सभी कैपिटल अक्षरों को भी छोटे अक्षरों से बदला जाता है।

स्टॉप वर्ड रिमूवल : यह शब्द सामान्यतः टेक्स्ट में उपयोग किए जाते हैं लेकिन दस्तावेज के विश्लेषण में महत्वपूर्ण नहीं है। अंग्रेजी स्टॉप शब्दों की सूची, जैसे पूर्ववर्ती, संयोजन, आलेख, क्रिया, संज्ञा, सर्वनाम, क्रिया, विशेषण, आदि। यह शब्द www-dcs-gla-ac-uk/idiom/ir_resources/linguistic_utils/stop_words से डाउनलोड किए जा सकते हैं। प्रयोज्य एक स्टॉप वर्ड लिस्ट के रूप में स्मार्ट डेटा सेट तैयार कर सकते हैं। उपयोगकर्ता परिभाषित सूची को स्टॉप शब्द के शब्दकोश के रूप में भी बनाया जा सकता है। जिसमें कुछ अन्य भाषा वर्ण, संख्याएं और अन्य विशेष वर्ण शामिल हैं।

स्टेमिंग : व्युत्पन्न शब्द को उनके आधार शब्द में परिवर्तित करने की प्रक्रिया को स्टेमिंग कहा जाता है। बेस शब्द को स्टेम के रूप में जाना जाता है जैसा कि हम जानते हैं कि प्रत्येक शब्द का प्रयोग कई रूपों में व्यक्त किया जा सकता है और विशेष अर्थ ले सकता है। उदाहरण के लिए, "कम्प्यूटरीकृत", "कम्प्यूटरीज", "कंप्यूटराइजेशन" और "कंप्यूटरशन" शब्द "कंप्यूटर"के रूप में समान आधार स्टेम साझा करते हैं। सभी को अलग-अलग शब्दों के रूप में माने जाने की बजाय, इसे एक शब्द के द्वारा प्रतिस्थापित किया जा सकता है। इस तरह, ऐसे पदों के लिए भार को बढ़ाया जा

सकता है और अनावश्यक शब्दों की संख्या से बचा जा सकता है। स्टैंडर्ड पोर्टर स्टॉमिंग एल्गोरिदम का उपयोग स्टेमिंग के लिए इस्तेमाल किया जा सकता है (पोर्टर, 1981)।

चरों का चुनाव : डेटासेट से अक्सर अधिक बार और कम आवृत्ति वाले शब्दों को हटाया जाता है वे दस्तावेजों के साथ भेदभाव करने में सक्षम नहीं हो सकते हैं या मॉडल को परिभाषित नहीं कर पाते हैं। आम तौर पर, वर्गीकरण के सन्दर्भ में, जो शब्द चालीस /पचास से अधिक और तीन या उससे कम हैं, उनको आम तौर पर निकाल दिया जाता है क्योंकि ज्यादातर डेटा माइनिंग एल्गोरिथ्म बड़े फीचर सेट को संभालने में सक्षम नहीं होते हैं।

इनफार्मेशन गेन या इन्फोगेन: इसका उपयोग महत्वपूर्ण चरों के चुनाव के लिए किया जाता है यहाँ पर एन्ट्रॉपी इनफार्मेशन गेन शब्द या टोकन के महत्व या प्रासंगिकता को निर्धारित करने में सहायक होता है। शीर्ष कुछ शब्दों को चुनना डेटा सेट में गैर-प्रासंगिक सुविधाओं को हटाने में मदद करता है। मशीन लर्निंग की तकनीक के प्रदर्शन में किसी भी महत्वपूर्ण सुधार को दिखाने के लिए शब्दों की संख्या के प्रभाव का निर्धारण करने में भी सहायक होते हैं।

टेक्स्ट डेटा का रिप्रजेंटेशन (प्रतिनिधित्व) :

टेक्स्ट डेटा आम तौर पर असंरचित प्रारूप में उपलब्ध है। इसे विश्लेषणात्मक के अनुकूल बनाने की आवश्यकता है। टेक्स्ट डेटा को कई विशेषताओं में विभाजित किया जा सकता है, जैसे शीर्षक, वर्ग, उपखंड, पैराग्राफ इत्यादि। टेक्स्ट की रिपोर्ट संरचित करने के कई तरीके हैं। प्रत्येक दस्तावेज में कई टोकन होते हैं और कई दस्तावेज भी उपलब्ध हैं। इसे (दस्तावेज X शब्द) को एक सारणी के रूप में दर्शाया जा सकता है जहाँ पंक्तियों को दस्तावेज और शब्दों को स्तंभ के रूप में माना जाता है, (Salton and Buckley,1988) इस सारणी के गुणों को निम्नलिखित तरीकों से परिभाषित किया जा सकता है।



बाइनरी रिप्रजेंटेशन (प्रतिनिधित्व) : सारणी का शून्य या एक के द्वारा दस्तावेज में संबन्धित पद की अनुपस्थिति या उपस्थिति के लिए लिया जाएगा। इस प्रकार का प्रतिनिधित्व संभवतः उपयुक्त बेस्ड मॉडल जैसे कि नेव बेयस (Naive Bayes) वर्गीकरण में उपयोगी होगा, जहां किसी पद की उपस्थिति या अनुपस्थिति पर विचार किया जाता है।

$$W_i = \begin{cases} 1, & \text{यदि शब्द दस्तावेज में उपलब्ध है} \\ 0, & \text{अन्यथा} \end{cases}$$

इस प्रकार का प्रतिनिधित्व एक बड़े दस्तावेज के प्रति पक्षपातपूर्ण है।

टर्म फ्रीक्वेंसी रिप्रजेंटेशन : इस प्रकार के प्रतिनिधित्व में, दस्तावेज में उपलब्ध शब्दों की आवृत्ति को गिना जाएगा और सारणी में संग्रहित किया जाता है। इस प्रकार का प्रतिनिधित्व एक दस्तावेज में किसी शब्द को विशेष महत्व देता है जिसे मल्टीनोमिअल नैव बेयस (Naive Bayes) एल्गोरिथम द्वारा वर्गीकरण के रूप में किया जाता है।

TF_i = शब्द की दस्तावेज में आवृत्ति

इस मानक को सामान्यीकृत किया जा सकता है, यदि हम दस्तावेज में कुल संख्या की पद के द्वारा टीएफ (टर्म फ्रीक्वेंसी) को विभाजित करते हैं तो

जहां दस्तावेज में कुल पदों संख्या की संख्या

$$Norm(TF_i) = \frac{TF_i}{n}$$

N है।

टीएफ X आईडीएफ ($TF \times IDF$) का रिप्रजेंटेशन (प्रतिनिधित्व) टीएफ X आईडीएफ "टर्म आवृत्ति (टीएफ) समय व्युत्क्रम दस्तावेज आवृत्ति (आईडीएफ)" है। यह आमतौर पर पूर्ण डाटासेट या दस्तावेज सेट में किसी शब्द के महत्व को निर्धारित करने के लिए उपयोग किया जाता है। व्युत्क्रम दस्तावेज आवृत्ति शब्द के साथ दस्तावेजों की संख्या से सभी दस्तावेजों की संख्या को विभाजित करके प्राप्त की जाती है, और फिर उसका लॉगरिथम लिया जाता है। इस तरह के प्रतिनिधित्व का इस्तेमाल शब्दों की रैंकिंग में किया जाता है और शीर्ष कुछ पदों का चयन किया जाता है।

$$W_i = TF_i * IDF_i$$

$$\text{जहां } IDF_i = \log \left(\frac{N}{DF_i} \right)$$

DF_i दस्तावेज आवृत्ति है और दस्तावेजों की संख्या के रूप में परिभाषित किया गया है। N कुल दस्तावेजों की संख्या है।

माइनिंग टेक्स्टुअल डाटा

एक बार जब यह सारणी तैयार हो जाती है तब डाटा माइनिंग की तकनीक वांछनीय कार्य के आधार पर लागू की जा सकती है। डेटा माइनिंग के समान, मौजूदा साहित्य के साथ नए ज्ञान की खोज और अनुमान लगाने के लिए सांख्यिकीय, डेटा माइनिंग, मशीन लर्निंग और सॉफ्ट कंप्यूटिंग तकनीकों का उपयोग करते हुए इस डेटा से महत्वपूर्ण जानकारी को निकालने के कई तरीके हैं। टेक्स्ट माइनिंग में विभिन्न कार्यों की श्रेणियां डेटा अन्वेषण, वर्गीकरण, क्लस्टरिंग, और विजुअलाइजेशन हैं। सामान्य आंकड़ों को डेटा के आवृत्ति वितरण और परिवर्तन के संबंध में कई तरीकों से देखा जा सकता है। नेचुरल लैंग्वेज प्रसंस्करण के स्तरों को सरणी 1 में दर्शाया गया है।

सारणी 1: एनएलपी के प्रसंस्करण स्तर

टेक्स्ट माइनिंग के उपयोग

प्रसंस्करण स्तर	कार्य और अनुप्रयोग
अक्षरस और स्ट्रिंग्स	शब्द टोकनकरण, वाक्य सीमा का पता लगाने, जीन प्रतीक मान्यता, टेक्स्ट पैटर्न निष्कर्षण
शब्द टोकन	पार्ट ऑफ स्पीच-टैगिंग, पद विच्छेदन, विभक्त करना, शब्द निष्कर्षण, जीन मान्यता का उल्लेख
वाक्य	वाक्य वर्गीकरण और पुनः प्राप्ति और श्रेणी, प्रश्न उत्तर, स्वतः संक्षिप्तीकरण
वाक्य विंडो	शब्दों को विभक्तिकरण
पैराग्राफ एंड पैसेज पूरा दस्तावेज	शब्दों और उससे जुड़े क्षेत्रों की जांच दस्तावेज समानता गणना
बहु दस्तावेज	डॉक्यूमेंट क्लस्टरिंग, बहुदस्तावेज सारांश



लाइन, उक्तक, परिस्थितियों) की प्रासंगिक जानकारी के रूप में सामने आ सकती है। बेहतर जानकारी के लिए उपयोगकर्ता की मांग (खोजशब्द खोजों से परे) जानकारी की तीव्र वृद्धि, मैनुअल सूचना निष्कर्षण कुशल नहीं है और वर्तमान समय में पूर्ण टेक्स्ट लेख, डेटाबेस और जीनोमिक अध्ययन से जानकारी का एकीकरण किया जा सकता है।

टेक्स्ट विश्लेषण और क्यूरेशन प्रक्रिया चित्र 2. जीन एनोटेशन का उपयोग करते हुए संशोधन के साथ मैनुअल क्यूरेशन, फीचर्स, एनोटेशन और पूर्वानुमान के साथ शुरू होती है



चित्र 2. टेक्स्ट विश्लेषण और क्यूरेशन की प्रक्रिया

टेक्स्ट माइनिंग में चुनौतियाँ

टेक्स्ट माइनिंग समुदाय द्वारा कई चुनौतियों का सामना किया जा रहा है। ज्ञान / जानकारी के तेजी से और निरंतर सामयिक बनने, असंरचित टेक्स्ट डेटा, साहित्य / दस्तावेजों में एक से अधिक भाषा का उपयोग, इस चुनौती को संभालने के लिए कोई भी नया ज्ञान और पैटर्न अनुमान लगाने में मुश्किल हो रही है। महत्वपूर्ण चुनौतियों में से कुछ नीचे दी जा रही हैं

- असंरचित शाब्दिक रूप में उपलब्ध जानकारी।
- कम्प्यूटर के द्वारा आसानी से उपयोग एवं समझने योग्य जानकारी
- दस्तावेजों के विशाल संग्रह के साथ लेनदेन

- समानार्थियों / विलोम शब्द / पीओएस टैगिंग आदि से निपटने में समस्याएं
- शब्दकोश की प्राथमिकता और प्राथमिक शब्द की पहचान
- अन्य कोडिंग या अन्य ओर मिश्रित भाषा वर्ण

टेक्स्ट माइनिंग के लिए उपकरण

टेक्स्ट माइनिंग करने के लिए बड़ी संख्या में उपकरण उपलब्ध हैं। ओपन सोर्स टूल्स और कमर्शियल टूल्स में निम्नानुसार हैं

उपकरण के ओपन सोर्स श्रेणी में टेक्स्ट और खोज परिणामों के क्लस्टरिंग फ्रेमवर्क, गेट (जनरल आर्किटेक्चर फॉर टेक्स्ट इंजीनियरिंग) के लिए इस्तेमाल की जाने वाली कैरट2 (Carrot2) में नेचुरल लैंग्वेज टूलकिट और नेचुरल इंजीनियरिंग के लिए ओपन सोर्स टूलबॉक्स, बड़े पैमाने पर विषय मॉडलिंग और निकासी के लिए पायथन (Python) आधारित जेन्सिम का उपयोग असंगठित टेक्स्ट से सिमेंटिक जानकारी, नेचुरल लैंग्वेज प्रोसेसिंग के लिए ओपनएनएलपीपी (OpenNLP), नेचुरल लैंग्वेज टूलकिट (NLTK), प्रतीकात्मक और सांख्यिकीय नेचुरल लैंग्वेज प्रोसेसिंग (एनएलपी) के लिए पुस्तकालयों और कार्यक्रमों का एक सूट के रूप में उपलब्ध है। ऑरेंज टेक्स्ट माइनिंग ऐंड-ऑन, R टेक्स्ट माइनिंग के टेक्स्ट प्रसंस्करण विस्तार के रूप में KNIME PLUS टेक्स्ट माइनिंग संग्रह आदि के लिए एक रूपरेखा प्रदान करता है।

उपकरणों की व्यावसायिक श्रेणी में आईबीएम एसपीएसएस – मॉडलर का प्रदाता (जिसे पहले आईबीएम एसपीएसएस मॉडलर और आईबीएम एसपीएसएस टेक्स्ट एनालिटिक्स कहा गया था) भी शामिल हैं एनएलपी आधारित टेक्स्ट विश्लेषण क्षमताओं (बहुभाषी भावना, घटना और तथ्य निष्कर्षण) का वर्गीकरण मैथमैटिका प्रदान करता है। डेटा संरक्षण, पैटर्न मेचिंग, क्लस्टरिंग और सिमेंटिक विश्लेषण के लिए टूल बनाया गया है। डेटा और टेक्स्ट माइनिंग के लिए अपनी टेक्स्ट प्रसंस्करण विस्तार के साथ रैपिड माइनर (Rapid Miner) चित्र 3. टेक्स्ट विश्लेषण के लिए रैपिड माइनर टूल के मुख्य पृष्ठ

पर उपलब्ध फीचर और फंक्शन को दर्शाया गया है, एसएस टेक्स्ट खनिक और वाणिज्यिक टेक्स्ट विश्लेषिकी, प्राकृतिक भाषा प्रसंस्करण, और सूचना प्रबंधन, STATISTICA के लिए उपयोग किए जाने वाले वर्णनात्मक विश्लेषिकी सॉल्यूशंस के लिए डेटा माइनर में एक वैकल्पिक विस्तार के रूप में टेक्स्ट माइनर, मेगा कंप्यूटर इंटेलिजेंस टेक्स्ट के बड़े संस्करणों और नेचुरल लैंग्वेज प्रोसेसिंग (NLP), मशीन सीखने, भावना विश्लेषण, इकाई निष्कर्षण, क्लस्टरिंग और वर्गीकरण सहित संरचित डेटा से कार्रवाई योग्य ज्ञान प्राप्त करता है। एमआईटी मीडिया लैब में नेचुरल लैंग्वेज प्रोसेसिंग (एनएलपी), मशीन लर्निंग और कृत्रिम बुद्धि अनुसंधान के लिए विकसित किया गया है। ये उपकरण नए ज्ञान और विचारों का अनुमान लगाने के लिए, बड़ी संख्या में रिपोर्ट, साहित्य, दस्तावेज, प्रतिक्रिया आदि सीखने, समझने और मापने में सहायक है।



चित्र 3. रैपिडमाइनर टूल

डेटा पूर्व प्रसंस्करण के वर्क फ्लो में (चित्र 4.) जहाँ टेक्स्ट दस्तावेज का टोकनकरण करते हुए शब्दों को उनके आधार शब्द में परिवर्तित करते हुए व स्टॉप वर्ड (शब्दों) को हटाते हुए एवं अन्य शब्दों को अलग करते हुए दर्शाया गया है

उदाहरण के रूप में W&J 48 मॉडल का चयन करने के पश्चात टेक्स्ट डाटा का परीक्षण करते हुए शब्दों का निष्कर्षण एवं इन्फोगेन का उपयोग करते हुए शर्तों का चयन करने के पश्चात परिणाम को सत्यापित करते हुए मॉडल को लागू किया जाता है, उसके बाद



चित्र 4. वर्क फ्लो डाटा पूर्व प्रसंस्करण

मॉडल का मूल्यांकन निष्पादित किया जाता है। चित्र 5 में क्लासिफायर बिल्डिंग्स के विभिन्न चरणों को दर्शाया गया है।



चित्र 5. क्लासिफायर बिल्डिंग्स

निष्कर्ष

टेक्स्ट माइनिंग/विश्लेषिकी साहित्य खनन से सोर्स कोड का विश्लेषण करने के लिए अलग-अलग प्रत्येक डोमेन में किया जाता है। ओपन सोर्स या व्यावसायिक उपकरणों की मदद से विभिन्न प्रकार के प्लेटफार्मों में डिजिटल सामग्रियों की उपलब्धता, इन बड़े दस्तावेजों और रिपोर्टों को संसाधित करने में कुछ आसान बनाती है। स्पैम फिल्टरिंग, जैविक साहित्य माइनिंग, सूचना पुनःप्राप्ति की तकनीकियों में पिछले कुछ वर्षों में काफी सुधार हुआ है। टेक्स्ट माइनिंग क्षेत्र, कृषि में अनुप्रयोग के क्षेत्रों की एक बड़ी संख्या



के साथ-साथ आगे बढ़ रहा है। अतः यह जैव सूचना विज्ञान के साहित्य में उपलब्ध जानकारी को खोजने में महत्वपूर्ण सिद्ध हो सकता है।

संदर्भ:

1. सल्टन एवं बकले (Salton and Buckley|1988) टर्म-वेइटिंग अप्रोचेस इन ऑटोमैटिक टेक्सट रिट्रीवल। इनफार्मेशन प्रोसेसिंग ऐंड मैनेजमेंट Vol- 24, No- 5, pp- 513&523
2. पोर्टर, एम.एफ. (Porter, M-F., 1980) एन अल्गोरिथम फॉर सफिक्स स्ट्रिपिंग। प्रोग्राम,14(3), pp-130-137
3. नेव बेयस (Naive Bayes) टेक्सट वर्गीकरण (<http://nlp-standford-edu/IR&book/html/html/edition/naive&bayes&teÜt&classification&1-html>)
4. स्टॉप वर्ड रिमूवल (www-dcs-gla-ac-uk/idom/ir_resources/linguistic_utils/stop_words)

फ़ज़्जी तकनीक से पशुधन निर्यात का विश्लेषण

रविन्द्र सिंह शेखावत, के. एन. सिंह, विशाल गुरुंग, राजीव रंजन कुमार,
अचल लामा एवं मो. यासीन

सार

भविष्य की योजना बनाने के लिए अर्थमिति में समय श्रृंखला डेटा का विश्लेषण महत्वपूर्ण भूमिका निभाता है। समय श्रृंखला डेटा के पूर्वानुमान के लिए उपयोग होने वाले पारंपरिक मॉडल एवं उपकरण के परिणाम क्रिप्स, नियतात्मक और सटीक होते हैं। अतः यह वास्तविक दुनिया के समय-श्रृंखला की समस्याओं में प्रयोग के लिए प्रतिबंधात्मक होता है। इसलिए, इस तरह की सीमाओं को दूर करने के लिए, एक शक्तिशाली फ़ज़्जी तंत्र को नियोजित किया गया है। इस लेख में, फ़ज़्जी तकनीक का उपयोग करके भारत के पशुधन निर्यात समय-श्रृंखला डेटा में उतार-चढ़ाव का अध्ययन किया गया है।

कुंजी शब्द: क्रिप्स वेरिएवल, फ़ज़्जी सेट सिद्धांत, संशोधित एफटीएस मॉडल।

परिचय

समय-श्रृंखला डेटा की मॉडलिंग एवं पूर्वानुमान के लिए अधिकांश पारंपरिक तकनीकें क्रिप्स, नियतात्मक एवं सटीक होती हैं। हमारा मुख्य उद्देश्य उन मॉडलों का उपयोग करना है जो निष्पक्ष और सटीक पूर्वानुमानित करने के लिए कुशल हैं और व्यावहारिक समय-श्रृंखला डेटासेट के कई महत्वपूर्ण विशेषताओं को हासिल करने में सक्षम हैं। इसलिए वास्तविकता का वर्णन करने के लिए अधिक विस्तृत अध्ययन की आवश्यकता है। इस संदर्भ में, फ़ज़्जी सेट सिद्धांत को सबसे पहले जडे और गोगुएन (1973) द्वारा प्रस्तावित किया गया था। टाइम-सीरीज डेटा की मॉडलिंग के फ़ज़्जी तकनीक में समय श्रृंखला मॉडल के घटकों को फ़ज़्जी सेट के रूप में माना जाता है। जब डेटा में अस्पष्टता और अदर्शन की उपस्थिति होती है तो फ़ज़्जी तकनीक का उपयोग करते हैं। फ़ज़्जी पद्धति को पहली बार गीत और

चिसोम (1993) द्वारा प्रस्तावित किया गया था। जैसा कि इस मॉडल में जटिल गणितीय विश्लेषण शामिल है, ली, लियू और चेन (2006) ने एक अधिक सरल लेकिन कुशल मॉडल का प्रस्ताव किया ताकि कार्यप्रणाली को मजबूत किया जा सके और सटीक पूर्वानुमान मूल्य प्राप्त कर सकें। इसके अलावा, चेन के मॉडल की कुछ सीमा को संबोधित करने के लिए एक संशोधित फ़ज़्जी समय श्रृंखला की भविष्यवाणी पद्धति का प्रस्ताव किया गया। इस लेख का मुख्य उद्देश्य फ़ज़्जी तकनीक का उपयोग करते हुए पशुधन निर्यात समय-श्रृंखला डेटा को मॉडल करना तथा ऑटोरेग्रेसिव इंटीग्रेटेड मूविंग एवरेज (ARIMA) मॉडल की परंपरागत पद्धति पर सुधार दिखाना है।

सामग्री एवं तरीके

क्रिप्स वेल्यु को फ़ज़्जी वेल्यु में परिवर्तित करने की प्रक्रिया को फ़ज़्जीफिकेशन कहते हैं। यह रूपांतरण मेम्बरशिप फंक्शन द्वारा प्रस्तुत किया जाता है। डीफ़ज़्जीफिकेशन, फ़ज़्जीफिकेशन की रिवर्स प्रक्रिया है। डीफ़ज़्जीफिकेशन में, फ़ज़्जी वेल्यु को क्रिप्स वेल्यु में परिवर्तित करते हैं। वेटेड औसत विधि, अधिकतम सदस्यता सिद्धांत, आदि डीफ़ज़्जीफिकेशन के लिए कई तरीके उपलब्ध हैं।

माना कि, y_t ($t=1,2,\dots,n$) समय श्रृंखला के वेरिएबल हैं, जो वास्तविक संख्या का एक सबसेट है। $f_t(i)$ ($i=1,2,\dots,n$) एक परिभाषित काल श्रृंखला पर फ़ज़्जी सेट है। तब $F(t)$, जो कि सभी $f_t(i)$ का संग्रह है, को फ़ज़्जी टाइम सीरीज कहते हैं।

फ़ज़्जी संबंध, फ़ज़्जी समय श्रृंखला के दो तत्वों के बीच संबंध है। $R(t-1,t)$, $t-1$ और t समय बीच फ़ज़्जी संबंध को दर्शाता है अगर

$$F(t)=F(t-1)*R(t-1,t) \text{ है}$$



जहाँ * एक औपरेटर है

$F(t)$ और $F(t-1)$ के संबंध को इस प्रकार प्रदर्शित किया जा सकता है,

$$F(t-1) = F(t)$$

यह एक सिंगल फ़ज़्जी संबंध है। N-ऑर्डर फ़ज़्जी संबंध को इस प्रकार प्रदर्शित किया जा सकता है,

$$F(t-n)...F(t-2)F(t-1) = F(t)$$

चेन ने एक सरलीकृत मॉडल प्रस्तावित किया जो जटिल परिचालन को समाप्त करता है और गणना करने को आसान बनाता है। इस पद्धति में केवल सरल अंकगणितीय संचालन शामिल हैं। फ़ज़्जी समय श्रृंखला के संशोधित संस्करण में, डीफ़जिफिकेशन ऑपरेटर $Y(t)$ को इस प्रकार परिभाषित किया गया है,

यहाँ, a_{t-i} समय $(t-1)$ पर वास्तविक मान को दर्शाता है w_t एक वेट है जो उपयुक्त तकनीक द्वारा निर्धारित किया जाता है। n , संबंध का क्रम है (यलोको एट। अल, 2013)।

$$Y(t) = \sum_{i=1}^n a_{t-i} w_i$$

परिणाम एवं चर्चा

पशुओं उत्पादों का निर्यात के आंकड़े में बड़े पैमाने पर वेरिएशन हैं। इस अध्ययन में सबसे पहले पारंपरिक ARIMA पद्धति का इस्तेमाल किया गया है और फिर दो फ़ज़्जी तकनीकें, चेन मॉडल एवं संशोधित एफटीएस मॉडल का उपयोग मॉडलिंग के लिए किया गया है। एमएसई और एमएपीई के मूल्यों की तुलना करके, हम आसानी से अनुमान लगा सकते हैं कि दो फ़ज़्जी मॉडल जैसे, चेन मॉडल और संशोधित एफटीएस मॉडल पारंपरिक एआरआईएए मॉडल की तुलना में बेहतर प्रदर्शन करता है। तीनों मॉडलों से प्राप्त वास्तविक और पूर्वानुमानित वेल्सु तालिका 1 में प्रस्तुत किए गए हैं।

तालिका 1: मॉडल के एमएसई और एमएपीई

संक्षिप्त में कह सकते हैं कि फ़ज़्जी तकनीक, पशुधन निर्यात डेटा सेट के मॉडलिंग और पूर्वानुमान के लिए ARIMA मॉडलिंग पद्धति से बेहतर है।

	ARIMA(2,1,1)	Chen's model	Modified FTS model
MAPE	86.31236	55.35077	43.32458
MSE	13048.37	31130.94	2058.764

निष्कर्ष

इस लेख में समय-श्रृंखला डेटा की फिटिंग के लिए फ़ज़्जी तकनीक का उपयोग के बारे में वर्णन किया गया है। एमएसई और एमएपीई के वेल्सु से, हम अनुमान लगा सकते हैं कि संशोधित फ़ज़्जी टाइम सीरीज मॉडल समय श्रृंखला के मानों की भविष्यवाणी करने के लिए सबसे अधिक कुशल है। जैसा कि हम देख सकते हैं कि दोनों फ़ज़्जी तकनीक परंपरागत ARIMA विधि से अधिक सटीक अनुमान देते हैं। यह सुझाव दिया जाता है कि, टाइम-सीरीज डेटा के मॉडलिंग और पूर्वानुमान के लिए, शोधकर्ताओं को ARIMA मॉडल के बजाय फ़ज़्जी मॉडल का उपयोग करना चाहिए। भविष्य की अनुसंधान समस्याओं के रूप में, फ़ज़्जी मॉडल के अन्य संस्करणों के उपयोग की संभावना का पता लगाया जा सकता है।

संदर्भ

- एग्रीओग्लू ई, अलदाग सी एच और योलकू यू (2013). फ़ज़्जी टाइम सीरीज फोरकास्टिंग विद ए नोवल हाइब्रिड एप्रोच कम्बार्निंग फ़ज़्जी सी-मीन एंड न्यूरल नेटवर्क। *एक्सपर्ट सिस्टम विद एप्लीकेशन* 40 (3): 854-857।
- सईद ए ई एम बी एवं दजाव डी डी. (2012). एप्लीकेशन ऑफ ट्रिपोजाइडल फ्यूजिफिकेशन एप्रोच एंड पार्टिकल स्वान ऑप्टिमाइजेशन इन फ़ज़्जी टाइम सीरीज फोरकास्टिंग। *प्रोसिडिंग ऑन द इंटरनेशनल कॉन्फ्रेंस ऑन आर्टिफिशियल इंटेलिजेंस (आईसीएआई)*।
- सिंह पी. (2015). अ ब्रिफ रिव्यू ऑफ मॉडलिंग एप्रोचेज वेस्ड ऑन फ़ज़्जी टाइम सीरीज। *इंटरनेशनल जर्नल ऑफ मशीन लर्निंग एंड साइबरनेटिक्स*, 1-24
- योलकू यू, अलदाग सी एच, इग्रीओग्लू ई एवं यूस्लू वी आर. (2013). टाइम-सीरीज फोरकास्टिंग विद ए नोवल फ़ज़्जी टाइम सीरीज एप्रोच: एन एकजाम्पल फॉर इस्तांबुल स्टॉक मार्केट। *जर्नल ऑफ स्टैटिस्टिकल कम्प्यूटेशन एंड सिम्यूलेशन*, 83(4): 599-612।

विषम चरों के लिए समष्टि योग की प्रागुक्ति

प्रदीप बसाक, हुकुम चन्द्र, कौस्तव आदित्य, मृन्मय राय, वन्दिता कुमारी,
सुशील कुमार एवं ऊषा जैन

सार

व्यू एवं सिटर (2001) द्वारा की गयी मॉडल कैलीब्रेशन अप्रोच सर्वेक्षण आकलन के लिए कैलीब्रेशन की एक सामान्य अप्रोच है जो आकलन स्थिति में पूर्ण सहायक सूचना का दक्षतापूर्वक प्रयोग कर सकती है। यह अप्रोच रैखिक (लीनियर) एवं अरैखिक (नॉन-लीनियर), दोनों मॉडलों को हैंडल कर सकती है। यह इस अन्तर्निहित अवधारणा पर आधारित है कि स्टडी चर एवं इसके प्रागुक्त मान जो संयोजित ट्रांसफोर्मर्ड मॉडल के माध्यम से प्राप्त होते हैं उनके बीच रैखिक संबंध है यहाँ तक की जब स्टडी चर एवं सहायक चर अरैखिक है। फिर भी, मॉडल कैलीब्रेशन अप्रोच में प्रयुक्त प्रागुक्त मान बैक ट्रांसफॉर्मेशन के कारण अभिनत है। परिणामस्वरूप, समष्टि योग एवं माध्य के लिए परिणामी आकलक भी अभिनत हैं। हमने प्रागुक्त मानों के बायस करेक्टेड वर्जन का प्रस्ताव दिया तथा मॉडल-आधारित मॉडल कैलीब्रेशन अप्रोच का प्रयोग करते हुए समष्टि के आकलन को परिभाषित किया। हमने मॉन्टे कार्लो सिमुलेशन अध्ययन का प्रयोग करते हुए मौजूदा आकलकों की तुलना में अपने प्रस्तावित आकलक के संवर्धित निष्पादन का प्रदर्शन किया।

मुख्य-शब्द : मॉडल कैलीब्रेशन; बैक ट्रांसफॉर्मेशन अभिनति; टेलर सिरीज रेखीकरण

प्रस्तावना

डेविल एवं सार्नडल (1992) द्वारा प्रस्तावित कैलीब्रेशन अप्रोच का अब सर्वेक्षण आकलन के लिए व्यापक रूप से प्रयोग हो रहा है। सहायक सूचना उपलब्ध होने पर इस विधि से सर्वेक्षण आकलन में सुधार होता है। इस अप्रोच में मूल सर्वेक्षण भारों को संशोधित किया जाता है ताकि सहायक चर के ज्ञात

समष्टि लक्षणों, वास्तव में कुल (अथवा माध्य), को प्रतिदर्श से उत्पन्न किया जा सके। फिर भी, कैलिब्रेटेड प्रतिदर्श भारण निहित रूप से यह माना जाता है कि सर्वेक्षण चर Y एवं सहायक चर X रैखिक रूप से सम्बन्धित हैं। अनेक कृषि, बिजनेस सर्वेक्षणों में आंकड़े आमतौर पर विषम होते हैं तथा सर्वेक्षण चर एवं सहायक चरों के बीच सम्बन्ध रेखीय नहीं हो सकता है। ऐसे आंकड़ों के लिए मूल पैमाने में X एवं Y के बीच सम्बन्ध लीनियर न हों, परन्तु ट्रांसफोर्मर्ड पैमाने में लीनियर हो सकते हैं, उदाहरणतया, लॉगरिथम (लॉग) स्केल। ऐसे मामलों में, Y की रैखिक अवधारणा पर आधारित सर्वेक्षण आकलन, Y के ट्रांसफोर्मर्ड वर्जन के लिए इसी प्रकार की रैखिक अवधारणा पर आधारित सर्वेक्षण आकलन की तुलना में अकुशल हो सकते हैं। कार्लबर्ग (2000a), चन्द्र एवं चैम्बर्स (2011) तथा उनमें दिये गए संदर्भ देखें। अतः, यदि अन्डरलाइंग मॉडल अरैखिक है, रैखिक अवधारणा के अन्तर्गत प्राप्त कैलीब्रेशन एस्टीमेटर अकुशल हो सकते हैं। ऐसे अरैखिक डाटा को हैंडल करने के लिए, व्यू एवं सिटर (2001) ने मॉडल कैलीब्रेशन अप्रोच की व्याख्या की जो एक सामान्य मॉडल के अन्तर्गत कैलीब्रेशन प्रक्रियाओं को जरनलाइज करता है। अरैखिक आंकड़ों के लिए यद्यपि मॉडल कैलीब्रेशन विधि समष्टि कुल के लिए दक्ष आकलक उपलब्ध कराती है परन्तु ये आकलक अभिनत होते हैं। यह अभिनति इस कारण होती है कि मॉडल कैलीब्रेशन अप्रोच बैक ट्रांसफॉर्मेशन अभिनत करेक्शन पर विचार नहीं करता।

हमने उस स्थिति तक मॉडल कैलीब्रेशन अप्रोच का पता लगाया जहाँ तक रैखिक मॉडल डाटा को नियंत्रित करते हुए लॉग पैमाने पर रखता है। व्यू एवं सिटर (2001) के विपरीत, सर्वेक्षण आकलन के प्रति



हमारा दृष्टिकोण मॉडल आधारित होगा तथा लॉग प्रसमान्य समष्टि आंकड़ों के आकलन के लिए हमने एक मॉडल-आधारित मॉडल कैलीब्रेशन की व्याख्या की, जो लॉग पैमाने पर रैखिक मॉडल का अनुसरण करता है। मॉडल कैलीब्रेशन के माध्यम से प्राप्त कैलीब्रेशन भारों को परिभाषित करते हुए यह प्रस्तावित अप्रोच बैंक ट्रांसफॉर्मेशन अभिनत के लिए उत्तरदायी है।

2. लॉग ट्रांसफॉर्म मॉडल के अन्तर्गत आकलन

हम यह मान लेते हैं कि विषम समष्टि एवं सीमांत वितरण के साथ सर्वेक्षण चर y तथा सहायक चर x पूरी तरह सकारात्मक हैं तथा इन चरों के बीच सम्बन्ध कच्चे अथवा मूल पैमाने में रैखिक नहीं हैं परन्तु हम यह मान लेते हैं कि ट्रांसफॉर्म पैमाने पर सम्बन्ध रैखिक हैं, विशेष रूप से लॉगरिथमिक (लॉग) अर्थात् $\log(y)$ एवं $z = \log(x)$ (अथवा कई बार x) रैखिक हैं। तब हम मॉडल को इस प्रकार लिख सकते हैं :

$$l_i = \log(y_i) = z_i' \beta + \varepsilon_i; i = 1, \dots, N \quad (1)$$

जहाँ $z_i' = (1, \log(x_{i1}), \log(x_{i2}), \dots, \log(x_{ip}))$,

$$\beta = (\beta_0, \beta_1, \dots, \beta_p)' \text{ एवं } \varepsilon_i \sim N(0, \sigma^2)$$

लॉग प्रसमान्य वितरण के गुणों का प्रयोग करते हुए हम यह देख सकते हैं कि

$$E(y_i | x_i) = E(e^{l_i} | x_i) = \exp(z_i' \beta + \sigma^2 / 2)$$

$$\text{Var}(y_i | x_i) = \exp(2z_i' \beta + \sigma^2) \{ \exp(\sigma^2) - 1 \}.$$

प्राचलों के आकलित मानों का प्रयोग करते हुए, सर्वेक्षण चरों के प्रागुक्त मान हैं

$$\hat{Y}_i = \hat{E}(y_i | x_i) = \exp(z_i' \hat{\beta} + \hat{\sigma}^2 / 2); i = 1, \dots, N$$

यहाँ हम देखते हैं कि

$$E(\hat{Y}_i | x_i) = E \left\{ \exp(z_i' \hat{\beta} + \hat{\sigma}^2 / 2) \right\} \neq E(Y_i | x_i)$$

$$= \exp(z_i' \beta + \sigma^2 / 2).$$

यह अंकित करता है कि प्रागुक्त मान अर्थात् $\hat{Y}_i = \exp(z_i' \hat{\beta} + \hat{\sigma}^2 / 2)$, $i = 1, 2, \dots, N$ अभिनत हैं। यह अभिनति बैंक ट्रांसफॉर्मेशन के कारण है। अतः, प्रागुक्त मानों पर बैंक ट्रांसफॉर्मेशन अभिनति करेक्शन टेलर

सीरीज एप्रोक्सीमेशन का निम्नानुसार प्रयोग करते हुए किया गया

हम यह मानते हैं कि $\hat{\eta} = (\hat{\beta}, \hat{\sigma}^2)'$, $\eta = (\beta, \sigma^2)'$ के प्रागुक्त मान है कि $E(\hat{\eta} - \eta) \approx 0$ बड़े n के लिए एवं $f(\eta) = \exp(z_i' \beta + \sigma^2 / 2)$ तब

$$f(\hat{\eta}) \approx f(\eta) + (\hat{\eta} - \eta)' f^{(1)}(\eta) + \frac{1}{2} (\hat{\eta} - \eta)' f^{(2)}(\eta) (\hat{\eta} - \eta)$$

$$\text{एवं } E\{f(\hat{\eta})\} \approx f(\eta) + \frac{1}{2} \text{tr} \left[E\{f^{(2)}(\eta) (\hat{\eta} - \eta) (\hat{\eta} - \eta)'\} \right]$$

टेलर सीरीज रैखीकरण को अपनाते हुए हम लिख सकते हैं

$$E\{f(\hat{\eta})\} \approx f(\eta) + \frac{1}{2} \exp(z_i' \beta + \sigma^2 / 2) \left[z_i' \text{var}(\hat{\beta}) z_i + \frac{1}{4} \text{var}(\hat{\sigma}^2) \right]$$

ताकि,

$$E(\hat{Y}_i) \approx \exp(z_i' \hat{\beta} + \hat{\sigma}^2 / 2) \left[1 + \frac{1}{2} \left\{ z_i' \text{var}(\hat{\beta}) z_i + \frac{1}{4} \text{var}(\hat{\sigma}^2) \right\} \right]$$

$$\approx k_i E(Y_i)$$

जहाँ $k_i = 1 + \frac{1}{2} \left\{ z_i' \text{var}(\hat{\beta}) z_i + \frac{1}{4} \text{var}(\hat{\sigma}^2) \right\}$ यही है, \hat{Y}_i, Y_i का अभिनत प्रागुक्त है। Y_i के बैंक ट्रांसफॉर्मेशन बायस करेक्टेड संयोजित मान $\hat{Y}_{i,BC} = \hat{k}_i^{-1} \hat{Y}_i$ द्वारा दिये गये हैं, जहाँ "BC" की सबस्क्रिप्ट का "बायस करेक्टेड" वर्जन दर्शाने के लिए प्रयोग किया गया है। मॉडल (1) के अन्तर्गत, हमारे पास है

$$\text{var}(\hat{\beta}) = \sigma^2 (Z_i' Z_i)^{-1} \text{ एवं } \text{var}(\hat{\sigma}^2) = \frac{2\sigma^4}{\{n - (p+1)\}}$$

जहाँ $Z_i = (z_1, z_2, \dots, z_n)'$, $Z_U = (z_1, z_2, \dots, z_N)'$ के प्रतिदर्श मानों की $n \times p$ मैट्रिक्स को दर्शाता है। इस प्रकार दिये गये मॉडल के अन्तर्गत, फॅक्टर k_i अर्थात् बायस करेक्शन निम्न द्वारा दिया गया है

$$k_i = 1 + \frac{1}{2} \left\{ \sigma^2 z_i' (Z_i' Z_i)^{-1} z_i + \frac{\sigma^4}{2\{n - (p+1)\}} \right\}$$

बायस करेक्शन का आकलन है

$$\hat{k}_i = 1 + \frac{s^2 a_i}{2} + \frac{s^4}{4n}, \text{ साथ में } a_i = z_i' (Z_i' Z_i)^{-1} z_i.$$

अब हम y_i एवं $\hat{y}_{i,BC}$ के बीच रैखीय मॉडल लिखेंगे

$$E(y_i | \hat{y}_{i,BC}) = \alpha_{0,BC} + \alpha_{1,BC} \hat{y}_{i,BC} \text{ एवं}$$

$$\text{Cov}(y_i, y_j | \hat{y}_{i,BC}, \hat{y}_{j,BC}) = \omega_{ij} \quad (2)$$

वैकल्पिक रूप से, मॉडल (2) को भी इस प्रकार लिखा जा सकता है

$$E(y_i | \hat{y}_{i,BC}) = \mu(j_{i,BC}, \theta) = j'_{i,BC} \theta \text{ एवं}$$

$$\Omega_U = [\omega_{ij}; i = 1, \dots, N; j = 1, \dots, N],$$

जहाँ $j'_{i,BC} = (1, \hat{y}_{i,BC})$ एवं $\theta = (\alpha_{0,BC}, \alpha_{1,BC})'$. हम ये मानते हैं कि $J_{U,BC}$ 'डिजाइन मैट्रिक्स' को दर्शाता है जो $J_{U,BC} = (j'_{1,BC}, \dots, j'_{N,BC})'$ के रूप में परिभाषित है।

अब हम प्राचलों के आकलित मानों का प्रयोग करते हैं (अर्थात् $\hat{\beta}$ and $\hat{\sigma}^2$) तथा $\hat{\Omega}_U$ एवं $J_{U,BC}$ के प्रतिदर्श एवं गैर-प्रतिदर्श विभाजनों पर विचार करेंगे। तब भार जो समष्टि कुल की EBLUP को परिभाषित करते हैं, इस प्रकार दिये गये हैं

$$w^{mc,BC} = (w_i^{mc,BC}; i \in S) =$$

$$I_s + \hat{H}'_{mc,BC} (J'_{U,BC} I_U - J'_{s,BC} I_s) \quad (3)$$

जहाँ $\hat{H}_{mc,BC} = (J'_{s,BC} \hat{\Omega}_U^{-1} J_{U,BC})^{-1} J'_{s,BC} \hat{\Omega}_U^{-1} \hat{\Omega}_U$

$$= \text{diag} [\exp(2z'_i \hat{\beta} + \hat{\sigma}^2) \{ \exp(\hat{\sigma}^2) - 1 \}; i = 1, \dots, n]$$

एवं I_U तथा I_s क्रमशः N एवं n आकार के 1's के सदिशों को दर्शाता है। ये 'बायस करेक्टेड संयोजित मान' मॉडल (2) के तहत Y की समष्टि कुल की प्रागुक्ति के लिए मॉडल-आधारित मॉडल-कैलीब्रेटेड भार है। चूंकि $J'_{s,BC} w^{mc,BC} = J'_{U,BC} I_U$ है, (2) के तहत ये भार (3) मॉडल-कैलीब्रेटेड हैं। यह कैलीब्रेशन प्रतिबन्धों को संतुष्ट करता है

$$\sum_{i \in S} w_i^{mc,BC} = N \text{ एवं } \sum_{i \in S} w_i^{mc,BC} \hat{y}_{i,BC} = \sum_{i \in U} \hat{y}_{i,BC}$$

$w^{mc,BC} = (w_i^{mc,BC}; i \in S)$ भारों का प्रयोग करते हुए, परिमित समष्टि कुल के प्रस्तावित प्रागुक्त निम्न द्वारा दिये गये हैं

$$\hat{T}_{MCBC} = \sum_{i \in S} w_i^{mc,BC} y_i \quad (4)$$

4. आनुभविक मूल्यांकन

इस भाग में (4) (\hat{T}_{MCBC} द्वारा निरूपित) में दिये गये समष्टि कुल के प्रस्तावित मॉडल कैलीब्रेशन आकलक के परिमित प्रतिदर्श के निष्पादन का मूल्यांकन तथा चार अन्य आकलकों, जैसे कि (i) सरल प्रतिदर्श माध्य (\hat{T}_R द्वारा निरूपित), (ii) रैखिक कैलीब्रेशन आकलक (\hat{T}_{LC} द्वारा निरूपित), (iii) कार्लबर्ग (2000a) के प्रागुक्ति एप्रोच आधारित एस्टीमेटर (\hat{T}_K द्वारा निरूपित) तथा बैक ट्रांसफॉर्मेशन बायस करेक्शन के बिना व्यू एवं सिटर के मॉडल कैलीब्रेशन एस्टीमेटर (\hat{T}_{MC} द्वारा निरूपित) से इनकी तुलना

करने के लिए हमने अनुकार अध्ययन संचालित किया।

अनुकरण में, मॉडल $\log(y) = \beta_0 + \beta_1 x + \varepsilon$, से $N=2000$ इकाइयों की एक परिमित समष्टि जनरेट की गयी, जहाँ $x \sim \text{Gamma}(1,1)$ एवं $\varepsilon \sim N(0, \sigma^2)$ हैं। हमने यह माना है कि $\beta_0 = \beta_1 = 1$ है। σ^2 के विभिन्न मानों के चयन द्वारा तीन विभिन्न परिमित समष्टियां जनरेट की गयी थीं जिससे $\log(y)$ एवं x के बीच सह-सम्बन्ध गुणांकों क्रमशः 0.8, 0.7 तथा 0.6 हैं। प्रत्येक नियत परिमित समष्टि के लिए, बिना प्रतिस्थापन के सरल यादृच्छिक प्रतिचयन द्वारा $n = 50, 100, 150, 200$ आकार का एक प्रतिदर्श लिया गया। प्रतिदर्श आंकड़ों एवं सभी संयोजित मानों का प्रयोग करते हुए $\hat{T}_R, \hat{T}_{LC}, \hat{T}_K, \hat{T}_{MC}$ एवं \hat{T}_{MCBC} आकलकों को कम्प्यूट किया गया। मोन्टो कार्लो अनुकरण $M=5000$ बार चलाया गया। विभिन्न आकलकों का निष्पादन सापेक्ष अभिनति (RB) एवं सापेक्ष मूल माध्य वर्ग त्रुटि द्वारा मापी गयी। सह-सम्बन्ध गुणांकों के तीन विभिन्न मानों तथा चार विभिन्न प्रतिदर्श आकारों के लिए विभिन्न आकलकों के प्रतिशत सापेक्ष मूल माध्य वर्ग त्रुटि एवं प्रतिशत सापेक्ष अभिनति की मान तालिका 1 में दिये गये हैं। विभिन्न आकलकों की तुलना करने के लिए परिणामों के 12 सेट्स हैं।



तालिका 1 में परिणाम दर्शाते हैं कि सह-सम्बन्ध गुणांक के सभी मानों एवं प्रतिदर्श आकारों के लिए प्रस्तावित आकलक \hat{T}_{MCBC} के लिए प्रतिशत सापेक्ष अभिनति के मान आमतौर पर कम हैं। इसके अतिरिक्त, हमने सहसम्बन्ध में गिरावट के साथ सभी आकलकों के लिए सापेक्ष अभिनति के मानों में वृद्धि नोट की। प्रस्तावित प्रागुक्त में सापेक्ष अभिनति उस समय अधिक पायी गयी जब सहसम्बन्ध उच्च था। एक महत्वपूर्ण परिणाम यह है कि अन्य आकलकों के विपरीत, सहसम्बन्ध गुणांक के दिये गये सभी मानों के लिए, जैसे जैसे प्रतिदर्श आकार छोटा होता जाता है प्रस्तावित प्रागुक्त की सापेक्ष अभिनति कम होती जाती है जो यह इंगित करता है कि छोटे प्रतिदर्शों के लिए भी जहाँ अभिनति की समस्या अधिक है, प्रस्तावित आकलक अच्छे परिणाम उपलब्ध कराते हैं।

प्रतिशत सापेक्ष मूल माध्य वर्ग त्रुटि के मामले में, यह प्रस्तावित आकलकों \hat{T}_{MCBC} के लिए न्यूनतम है। यह भी देखा गया कि जैसे-जैसे ट्रांसफार्मर्ड चरों के बीच सहसम्बन्ध गुणांक घटता है, समस्त प्रतिदर्श आकारों के लिए प्रस्तावित आकलक की सापेक्ष मूल माध्य वर्ग त्रुटि में वृद्धि होती है। इसी प्रकार, जैसे-जैसे सहसम्बन्ध गुणांक के दिये गये सभी मानों के लिए प्रतिदर्श आकार छोटे होते जाते हैं वैसे-वैसे सापेक्ष मूल माध्य वर्ग त्रुटि में भी वृद्धि होती है। फिर भी, समान्य मॉडल कैलीब्रेशन की तुलना में सभी मामलों में बायस करेक्शन सहित मॉडल कैलीब्रेशन की सापेक्ष अभिनति एवं सापेक्ष मूल माध्य वर्ग त्रुटि में वृद्धि पर्याप्त है। कुल मिलाकर, सापेक्ष अभिनति एवं सापेक्ष मूल माध्य वर्ग त्रुटि दोनों के संदर्भ में प्रस्तावित प्रागुक्ति का निष्पादन सबसे उत्तम पाया गया। इसलिए विषम डाटा के लिए प्रस्तावित प्रागुक्ति का उपयोग बेहतर होगा।

तालिका-1: विभिन्न आकलकों की प्रतिशत सापेक्ष अभिनति (RB, %) एवं प्रतिशत सापेक्ष मूल माध्य वर्ग त्रुटि (RRMSE, %)

$\rho=0.8$					
आकलक	\hat{T}_R	\hat{T}_{LC}	\hat{T}_K	\hat{T}_{MC}	\hat{T}_{MCBC}
n=200 RB	1.754	-5.690	2.231	3.776	0.850
RRMSE	99.737	87.905	24.438	26.122	23.355
n=150 RB	-2.143	-9.639	2.357	4.579	0.421
RRMSE	110.569	100.327	26.294	28.768	24.880
n=100 RB	2.842	-10.046	2.741	6.451	0.135
RRMSE	148.056	123.307	30.624	34.729	28.915
n=50 RB	8.350	-16.093	2.530	11.071	-2.243
RRMSE	256.892	158.905	41.352	55.744	37.789
$\rho=0.7$					
n=200 RB	1.831	-5.571	3.777	6.547	1.524
RRMSE	100.687	88.723	34.609	38.690	32.049
n=150 RB	-1.958	-9.317	4.011	8.013	0.923
RRMSE	111.630	101.157	36.554	42.708	33.526
n=100 RB	2.922	-9.818	4.447	11.131	0.473
RRMSE	149.212	125.149	42.468	52.572	39.014
n=50 RB	8.394	-15.640	3.839	20.083	-2.415
RRMSE	257.286	159.500	57.163	96.321	51.960
$\rho=0.6$					
n=200 RB	1.898	-5.417	6.073	10.857	2.645
RRMSE	101.874	89.865	49.564	59.244	44.021
n=150 RB	-1.741	-8.920	6.439	13.407	1.889
RRMSE	112.909	102.102	50.848	65.802	45.100
n=100 RB	3.027	-9.470	6.857	18.586	1.394
RRMSE	150.929	127.846	58.808	83.570	53.026
n=50 RB	8.378	-15.071	5.491	36.159	-1.330
RRMSE	257.399	160.144	78.052	185.747	74.816



5. निष्कर्ष

कैलीब्रेशन के लिए प्रस्तावित मॉडल—आधारित मॉडल कैलीब्रेशन आकलक संयोजित मानों के बायस करेक्टेड वर्जन का उपयोग करता है यह विशेष रूप से तब प्रभावी होता है जब समष्टि विषम होती है। इसमें लागरिथ्मिक पैमाने पर एक रैखिक मॉडल संयोजित कर संयोजित मान प्राप्त करना और इस तरह बायस करेक्टेड संयोजित मान प्राप्त करना तथा तब सामान्य संयोजित मानों की अपेक्षा बैंक ट्रांसफॉर्मेशन बायस करेक्टेड संयोजित मानों पर कैलीब्रेटिंग शामिल है। इसके अतिरिक्त, इसमें माध्य वर्ग त्रुटि आकलन का लाभ है क्योंकि इस आकलक में भारित रैखीय फार्म है।

संदर्भ

घन्द्र, एच. एवं चैम्बर्स, आर.एल. (2011). स्मॉल एरिया एस्टीमेशन अण्डर ट्रांसफॉर्मेशन टू लीनियरिटी। *सर्वे मेथेडॉलॉजी*, 37, 39–51

डेविल, जे.सी. एवं सार्नडल, सी.ई. (1992). कैलीब्रेशन एस्टीमेटर्स इन सर्वे सैम्पलिंग। *जर्नल ऑफ*

दि अमेरिकन स्टैटिस्टिकल एसोसिएशन, 87, 376–382

कार्लबर्ग, एफ. (2000a) पॉपुलेशन टोटल प्रीडिक्शन अण्डर ए लॉगनार्मल सुपरपॉपुलेशन मॉडल। *मेट्रोन*, LVIII, 53–80

कार्लबर्ग, एफ. (2000b) सर्वे एस्टीमेशन फार हाईली स्क्यूड पॉपुलेशनस इन दा प्रेसेन्स ऑफ जीरोज। *जर्नल ऑफ आफिशियल स्टैटिस्टिक्स*, 16, 229–241

रॉयल, आर.एम. (1976). दि लीनियर लीस्ट-स्क्वेयर्स प्रीडिक्शन अप्रोच टू टू-स्टेज सैम्पलिंग। *जर्नल ऑफ दि अमेरिकन स्टैटिस्टिकल एसोसिएशन*, 71, 657–664

व्यु, सी. एवं सितर, आर.आर. (2001). ए मॉडल कैलीब्रेशन अप्रोच टू यूजिंग कम्प्लीट ऑकजीलरी इन्फॉर्मेशन फॉम सर्वे डाटा। *जर्नल ऑफ दि अमेरिकन स्टैटिस्टिकल एसोसिएशन*, 96, 185–193

स्वतंत्रता प्राप्ति के बाद भारत में चावल के उत्पादन की स्थिति एवं पूर्वानुमान

वसी आलम, रंजीत के. पॉल, प्रवीण आर्या एवं ऊषा जैन

चावल विकासशील विश्व की एक बहुत महत्वपूर्ण खाद्य फसल होने के साथ-साथ भारत की आबादी के 60 प्रतिशत से अधिक के लिए मुख्य भोजन है। भारत में, चावल के उत्पादन के संयोजित विकास में 1981-90 के दौरान 3.55 प्रतिशत से 1991-2000 में 1.74 प्रतिशत की गिरावट आई है (आध्या इत्यदि, 2001)। इस तथ्य के बावजूद कि भारत ने खाद्यान्न उत्पादन में सन्तोषजनक उपलब्धि प्राप्त की है परन्तु इसकी बढ़ती हुई आबादी ने इसे निष्प्रभावी कर दिया है। खाद्य एवं कृषि संगठन (एफ.ए.ओ.) का पूर्वानुमान है कि वैश्विक खाद्य उत्पादन में 2030 तक 40 प्रतिशत से अधिक तथा 2050 तक 70 प्रतिशत तक की वृद्धि की आवश्यकता होगी (एफ.ए.ओ., 2009)। वर्ष 1950 से 2011 तक भारत में चावल के उत्पादन, क्षेत्रफल एवं उपज के विस्तृत अध्ययन के लिए आर्थिकी एवं सांख्यिकी निदेशालय से आँकड़े प्राप्त किये गए हैं। सिंचित एवं गैर-सिंचित क्षेत्र की वजह से हुए अंशदायी उत्पादन के साथ-साथ भारत में चावल के उत्पादन की पद्धति का अध्ययन करने का प्रयास किया गया है। भारत में, गत 60 वर्षों में, चावल की उपज में 3.59 गुना, क्षेत्रफल में 1.43 गुना, उत्पादन में 5.12 गुना तथा सिंचित क्षेत्र में (प्रतिशत सिंचित क्षेत्र को फसलाधीन क्षेत्र से गुना कर आकलित) 2.72 गुना वृद्धि हुई है। प्रौद्योगिकीय प्रभाव के कारण हुई उपज के अलावा सिंचित क्षेत्र उत्पादन का एक प्रमुख निर्धारक है।

फसलीय क्षेत्र के साथ उत्पादन उपज का एक समान्य वस्तुनिष्ठ गुणक है। यदि उपज एवं फसलीय क्षेत्र, दोनों में वृद्धि होती है तो उत्पादन में भी वृद्धि होती है। दुर्भाग्य से, स्वतंत्रता प्राप्ति से अब तक, चावल के फसलीय क्षेत्र में केवल 1.43 गुना ही वृद्धि हुई है) मान लें कि, यदि फसलीय क्षेत्र में 1950 के क्षेत्र

से 2.00 गुना (अर्थात् 61.62 मि.हे.) की वृद्धि होती है, तो 2.4 ट./हे. वर्तमान उपज से अनुमानित उत्पादन 147.89 मि.ट. बनता है, जो वर्तमान फसलीय क्षेत्र (44.61 मि.हे.) से 3.35 ट./हे. उपज से प्राप्त उत्पादन के समकक्ष है। उपज सम्बन्धी उपलब्ध काल श्रृंखला आँकड़ों के अनुसार, उत्पादन के इतने बड़े लक्ष्य को प्राप्त करने के लिए फसलीय क्षेत्र (सिंचित) को बढ़ाने की अपेक्षा उपज में वृद्धि करना अधिक कठिन है। आबादी की बढ़ती हुई माँग को पूरा करने के लिए यह बेहतर विकल्प है।

चावल का उत्पादन वर्ष 1950 में 20.58 मिलियन टन (मि.ट.) की तुलना में वर्ष 2011 में 105.3 मिलियन टन (मि.ट.) था, जिससे उत्पादन अन्तर 84.72 मिलियन टन था। प्रथम दृष्टि में यह एक महत्वपूर्ण अन्तर दिखाई पड़ता है, वास्तव में, यह महत्वपूर्ण अन्तर न केवल प्रौद्योगिकीय प्रभाव (उच्च उपज किस्में, उर्वरक, समेकित नाशीजीव प्रबन्धन इत्यादि) के कारण है बल्कि फसलीय क्षेत्र में वृद्धि के कारण भी है। वर्ष 1950 में, चावल के अन्तर्गत फसलीय क्षेत्र 30.81 मि. हे. था, जो वर्ष 2011 में 44.01 मि.हे. बढ़ गया, गत 61 वर्षों में फसलीय क्षेत्र में अन्तर, अर्थात् 13.2 मि.हे. (44.01-30.81)।

वर्ष 1950 में, भारत में चावल की औसत उपज 0.668 ट./हे. थी, जो वर्ष 2011 में बढ़कर 2.4 ट./हे. हो गयी थी। यदि हम वर्ष 2011 की उपज लेते हैं, जो 2.4 ट./हे. थी (जिसपर 2011 के लिए उत्पादन 105.3 मिलियन टन आकलित किया गया है), 1950 के क्षेत्रफल के आधार पर, उत्पादन 20.58 मिलियन टन के स्थान पर 73.94 मिलियन टन बनता है तथा उत्पादन का अन्तर 31.36 मिलियन टन (105.3-73.94) बनता है और यह अन्तर मुख्यतः फसलीय क्षेत्र में



अन्तर के कारण है अर्थात् गत 61 वर्षों में 13.2 मि.हे (44.01-30.81)। चावल मानसून आधारित फसल है। असामयिक मानसून का आना तथा अत्यधिक वर्षा अथवा वर्षा में कमी का चावल के उत्पादन पर प्रतिकूल प्रभाव पड़ता है। चावल के उत्पादकों के लिए जोखिम को कम करने के लिए सिंचित क्षेत्र को बढ़ाने की आवश्यकता है।

सिंचित क्षेत्र पर एक नजर, चावल के सिंचित प्रतिशत क्षेत्र में 31.7 (1950 में) से 56.9 (2007 में) वृद्धि हुई है। सिंचित प्रतिशत क्षेत्र एवं कुल फसलीय क्षेत्र पर उपलब्ध काल श्रृंखला आँकड़ों के आधार पर, हमने सिंचित एवं गैर-सिंचित क्षेत्र के कारण प्रतिशत उत्पादन पर काल श्रृंखला आँकड़े जनरेट किये हैं तथा यह आरेख-1 में दर्शाया गया है। आरेख-1 से स्पष्ट है कि गैर-सिंचित क्षेत्र के कारण 1950 (68.3 प्रतिशत) से 2007 (43.1 प्रतिशत) उत्पादन में लगातार गिरावट आ रही है। इसके विपरीत, सिंचित क्षेत्र के कारण 1950 (31.7 प्रतिशत) से 2007 (56.9 प्रतिशत) तक उत्पादन में लगातार वृद्धि हो रही है। हम यह निष्कर्ष निकालते हैं कि चावल के उत्पादन के जो भी आँकड़े हैं वे केवल प्रौद्योगिकीय प्रभाव के कारण नहीं हैं बल्कि फसलीय क्षेत्र के कारण भी हैं। चावल के अन्तर्गत सिंचित क्षेत्र में वृद्धि की आवश्यकता है। चावल के उत्पादन के काल-श्रृंखला आँकड़ों की जाँच की गयी तथा ऑटो-रिग्रेसिव इण्टीग्रेटेड मूविंग एवरेज (बॉक्स, इत्यादि 1994) मॉडल के लिए उपयुक्त पाये गये। जैसा की हम जानते हैं कि ARMA (p,q)

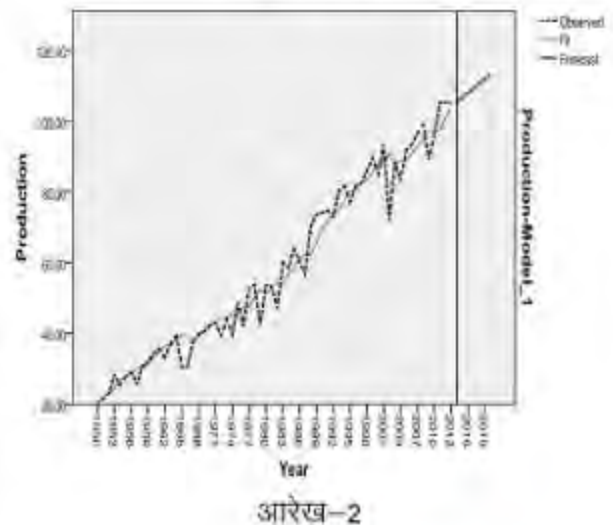
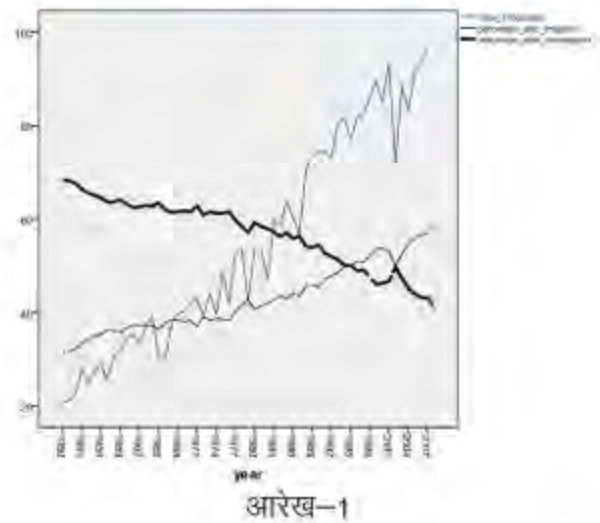
$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q} + \epsilon_t$$

जहाँ ϵ_t वाइट नॉइज है जो कि आई. आई. डि. है जिसका औसत = 0 है। सेरीज सटेशनरी नहीं होने के कारण प्रथम डिफरेंसिंग पर आँकड़ों को स्थायी बनाया गया तथा कुछ कैंडीडेट मॉडलों, जैसे ARIMA (1,1,0), ARIMA(1,1,1), और ARIMA (0,1,1) पर ट्राई किया गया। हमें लघुत्तम समान्यीकृत बी.आई.सी. अर्थात् ARIMA (1,1,1) एवं ARIMA (1,1,1) के लिए क्रमशः 3.408 तथा 3.487 की तुलना में ARIMA (0,1,1) के लिए 3.335 ज्ञात हुआ। अतः

ARIMA (0,1,1) मॉडल का उपयुक्त मॉडल के रूप में चयन किया गया, इस मॉडल के लिए प्राचल आकलक तालिका-1 में तथा पूर्वानुमान मान तालिका-2 में दिये गये हैं। एरिमा (0,1,1) का प्रयोग करते हुए आरेख-2 में वर्ष 2020 तक भारत में चावल के उत्पादन के पूर्वानुमान के साथ-साथ संयोजित मान दर्शाए गये हैं। वर्ष 2020 में पूर्वानुमान के अनुसार यह उत्पादन 113.26 मि. टन होगा।

तालिका-1 : एरिमा (0,1,1) मॉडल प्राचल

	आकलक	एस.ई	T	Sig
Constant	1.325	.172	7.686	.000
MA Lag1	.737	.090	8.172	.000





तालिका-2: पूर्वानुमान

मॉडल		2014	2015	2016	2017	2018	2019	2020
एरिमा (0,1,1)	पूर्वानुमान	105.31	106.63	107.96	109.28	110.61	111.93	113.26
	UCL	115.23	116.89	118.54	120.18	121.81	123.43	125.05
	LCL	95.39	96.38	97.38	98.39	99.41	100.43	101.46

संदर्भ

1. आध्या, टी.के., कार, पी. एवं सिन्हा, पी.के. (2011): विजन 2030, केन्द्रीय चावल अनुसंधान संस्थान, भारतीय कृषि अनुसंधान परिषद्, कटक (ओडीशा)।
2. बॉक्स, जी.ई.पी., जैनकिन्स, जी.एम. एवं रिन्सेल, जी.सी. (1994): टाइम सीरीस एनालिसिस : फोरकॉस्टिंग कन्ट्रोल, पीअर्सन एजुकेशन, दिल्ली।
3. एफ.ए.ओ. खाद्य एवं कृषि संगठन (2009), : ओ. ई.सी.डी.-एफ.ए.ओ. एग्रिकल्चरल आउटलुक 2009-2018

विकसित प्रतिमान का आकलन करने के लिए प्रसंभाव्य विभिन्नात्मक समीकरण पद्धति

हिमाद्रि घोष, सविता वधवा एवं प्रज्ञेषु

सार

जानवरों की विभिन्न प्रजातियों के विकास का वर्णन करने के लिए पशुविज्ञान में अरैखिक विकसित प्रतिमान व्यापक रूप से प्रयोग किए जाते हैं। अरैखिक पद्धतियों का उपयोग प्राचलों के आकलन के लिए किया जाता है। अरैखिक पद्धतियों की प्रथम सीमा यह है कि ये पद्धतियाँ तभी लागू हो सकती हैं, जब आँकड़े समसामयिक हों। दूसरी सीमा यह है कि नियोजित (deterministic) प्रतिमान में त्रुटि जोड़कर विशेष रूप से अनुदैर्घ्य आँकड़ों (Longitudinal) के लिए उनके अंतर्निहित उतार-चढ़ाव का वर्णन करने में उपरोक्त पद्धतियाँ सक्षम नहीं हैं। इस लेख का उद्देश्य पशु वैज्ञानिकों को एक नई पद्धति प्रसंभाव्य विभिन्नात्मक समीकरण (Stochastic Differential Equation) प्रतिमान से अवगत कराना है, जिसमें उपरोक्त सीमाओं का प्रतिबंधन नहीं है। गॉम्पर्टज़ विकसित प्रतिमान का उपयोग कर पद्धति का वर्णन किया गया। उपयुक्त एसएसएस (SAS) कोड विकसित किए गए। अंत में इस पद्धति का प्रयोग भारत के भारतीय पशु चिकित्सा अनुसंधान संस्थान, इज्जतनगर, बरेली के शूकर-बाड़े में एकत्रित शूकरों के मासिक वजन आँकड़ों पर किया गया।

प्रस्तावना

अरैखिक विकसित प्रतिमान जैसे, गॉम्पर्टज़, लॉजिस्टिक और वॉन बर्टलैनफी प्रतिमान पशु विज्ञान में पशुओं की विभिन्न प्रजातियों के विकास का वर्णन करने के लिए व्यापक रूप से प्रयोग किए जाते हैं। इन प्रतिमानों का महत्वपूर्ण पहलू यह है कि ये क्रियाविधिक प्रकृति के हैं और इसलिए अंतर्निहित प्राचलों की विशिष्ट जैविक व्याख्याएँ हैं। उदाहरणतः गॉम्पर्टज़ प्रतिमान के

तीन प्राचल आंतरिक वृद्धि दर, क्षमता और प्रतिक्रिया चर का प्रारंभिक आकार हैं। विकसित प्रतिमान अधिकतर अरैखिक विभिन्नात्मक समीकरण के रूप में व्यक्त किए जाते हैं। इन प्रतिमानों की विशेषता यह है कि इन्हें कुछ परिवर्तनों के माध्यम से रैखिक रूप में परिवर्तित किया जा सकता है, जैसे लघुगणकीय और व्युत्क्रम परिवर्तन। अंतर्निहित विभिन्नात्मक समीकरण के सटीक परिणाम प्राप्त किए जा सकते हैं जो कि प्राचल में अरैखिक होते हैं। इन्हें आँकड़ों में प्रयोग करने के लिए निर्धारित प्रतिमान में उचित मान्यताओं के साथ दाहिनी तरफ एडिटिव पद जोड़ दिया जाता है और अरैखिक पद्धति लैवनबर्ग-मार्कड (सैंबर और वाइल्ड, 2003) का प्रयोग करते हुए प्राचलों का आकलन किया जाता है। पिछले दो दशकों में इस पद्धति का वर्णन करने वाले बहुत से लेख अनुसंधान पत्रिकाओं में प्रकाशित हुए हैं; जैसे (वेणुगोपालन और प्रज्ञेषु 1997, प्रज्ञेषु और कंडाला 2003, प्रज्ञेषु और रविचन्द्रन 2003, मैटिस इत्यादि 2009 और घोष इत्यादि 2011)।

यद्यपि उपरोक्त पद्धति ने अतीत में कई उद्देश्यों को पूरा किया है परन्तु इसकी कुछ सीमाएँ हैं। प्रथम यह है कि यह पद्धति तभी लागू हो सकती है जब आँकड़े समसामयिक हों। उपरोक्त विकसित प्रतिमान में विकास दर प्रारंभिक चरण में तेज होती है और अगले चरण में धीमी हो जाती है जिससे सिगमोइड प्रकार का वक्र बनता है। इसलिए पशु वैज्ञानिक प्रारंभिक चरण में कम अंतराल पर और बाद के चरण में विस्तृत अंतराल पर वृद्धि आँकड़ों का अभिलेखन करते हैं। इसलिए असमान अंतरालों पर आँकड़े जनरेट हो जाते हैं। दूसरी सीमा यह है कि नियोजित प्रतिमान में त्रुटि जोड़कर विशेष रूप से अनुदैर्घ्य आँकड़ों (Longitudinal) के लिए उसके अंतर्निहित उतार-चढ़ाव का वर्णन करने में



उपरोक्त पद्धति (सैबर और वाइल्ड, 2003) सक्षम नहीं है। प्रसंभाव्य प्रक्रिया विकसित प्रतिमान को सीमाओं की बाध्यता से मुक्त करती है।

उपरोक्त सीमाओं को प्रसंभाव्य विभिन्नात्मक समीकरण (एसडीई) पद्धति (ओक्सेंडल, 2003) द्वारा नियोजित किया जा सका है। इसे विकसित नियोजित प्रतिमान के विभिन्नात्मक समीकरण में दाहिनी तरफ प्रसंभाव्य पद जोड़कर प्राप्त किया जा सकता है। ऐसा देखा गया है कि भौतिक स्थिति में जनसंख्या वृद्धि के समय-स्तर की तुलना में सामान्य रूप से जन्म और मृत्यु दर जैसे प्राचलों में भिन्नता के कारण यादृच्छिक पर्यावरण में उतार-चढ़ाव तेजी से होते हैं। इसलिए प्रसंभाव्य प्रक्रिया पद को गॉऊशियन व्हाइट नॉयज़ प्रसंभाव्य प्रक्रिया माना जाता है। परिणामस्वरूप प्रक्रिया मार्कोवियन हो जाती है जो कि एक अच्छा परिणाम है, हालांकि सैम्पल पाथ अनियमित है और पारंपरिक अर्थों में डेरिवेटिव को समायोजित नहीं करते हैं। इस स्थिति के लिए दो प्रकार के प्रसंभाव्य परिकलन क्रमशः स्ट्रैटोनोविच (Stratonovich) और ईटो (Itô) साहित्य में विकसित किए गए हैं। पहले परिकलन में सामान्य नियम लागू होते हैं जबकि दूसरे में संशोधित नियम लागू होते हैं। हालांकि इस लेख में एडिटिव नॉयज़ प्रतिमान का प्रयोग होने के कारण दोनों परिकलनों के परिणाम समान हैं। जिसमें स्टेट चर नहीं है (कोहेन और इलियट, 2015)। इस लेख में गॉम्पर्टज़ प्रतिमान के एसडीई संस्करण का अध्ययन किया गया है।

सामग्र्य और तरीके

गॉम्पर्टज़ विकसित प्रतिमान का विभिन्नात्मक समीकरण (सैबर और वाइल्ड 2003) निम्न है:

$$dy_t/dt = ry_t \log_e(K/y_t) \quad (1)$$

लघुगणकीय परिवर्तन से निम्न समीकरण प्राप्त हुआ

$$y_t = K \exp[\log_e(y_0/K) \exp(-rt)] \quad (2)$$

इससे संबंधित गॉम्पर्टज़ अरैखिक सांख्यिकीय विकसित प्रतिमान है:

$$y_t = K \exp[\log_e(y_0/K) \exp(-rt)] + \varepsilon_t \quad (3)$$

जहाँ ε_t स्वतंत्र और समान रूप से वितरित है, माध्य शून्य और परिसरण σ^2 है। यह तार्किक है कि

y_t में त्रुटि $\mu_t = K \exp[\log_e(y_0/K) \exp(-rt)]$ के समानुपातिक है अतः परिसरण स्थिरीकरण के लिए लघुगणकीय परिवर्तन उपयुक्त परिवर्तत है जो प्राचलों के आकलन के लिए आवश्यक है। तदनुसार परिवर्तित गॉम्पर्टज़ सांख्यिकीय विकसित प्रतिमान है:

$$\log_e y_t = [\log_e K + \log_e (y_0/K) \exp(-rt)] + \varepsilon_t^* \quad (4)$$

फीलिपे इत्यादि (2013) का अनुकरण करते हुए निरंतर प्रसार गुणांक के साथ अनुरूप गॉम्पर्टज़ एसडीआई प्रतिमान है:

$$dZ_t = r(\alpha - Z_t)dt + \sigma dW_t \quad (5)$$

जहाँ $\alpha = \log_e K$, $Z_t = \log_e y_t$ और W_t वीनर (Wiener) प्रक्रिया है जिसका परिसरण प्राचल एक है। समीकरण (5) के समरूप है:

$$d \exp(rt) Z_t = r \alpha dt + \sigma dW_t, \quad Z_{t_0} = z_0$$

दोनों तरफ एकीकरण के बाद और ईटो परिकलन का प्रयोग कर एसडीई प्रतिमान है:

$$Z_t = \alpha + (Z_{t_k} - \alpha) e^{-r(t-t_k)} + \sigma \exp(-rt) \int_{t_k}^t \exp(rs) dW_s \quad (6)$$

समीकरण (6) द्वारा दिए गए गॉम्पर्टज़ प्रतिमान का समाधान गॉशियन प्रक्रिया है, जिसका प्रतिबंधित माध्य एवं परिसरण है:

$$\begin{aligned} \mu_{Z_t|t_k} &= E\{Z_t|F_{t_k}\} = \alpha + (Z_{t_k} - \alpha) e^{-r(t-t_k)} \\ \sigma_{Z_t|t_k}^2 &= V\{y_t|F_{t_k}\} = \frac{\sigma^2(1 - e^{-2r(t-t_k)})}{2r} \end{aligned} \quad (7)$$

$\{Z_t, t \geq 0\}$ प्रक्रिया का ट्रांजिशन संभाव्यता धनत्व फलन माध्य और परिसरण क्रमशः α और $\sigma^2/(2r)$ के साथ उपगामी (Asymptotic) रूप से स्थिर है। जबकि $\{Z_t\}$ का माध्य-मान फलन अर्थात् $E\{Z_t\} = \alpha + (Z_{t_0} - \alpha) e^{-r(t-t_0)}$ सिगमाइड वक्र है। यह इंगित करता है कि समीकरण (6) द्वारा दिया गया समाधान त्रुटि प्रक्रिया $\{\sigma \exp(-rt) \int_{t_k}^t \exp(rs) dW_s, t \geq 0\}$ के साथ प्रतिमान विकसित करने में सक्षम है। अंत में गॉशियन यादृच्छिक चर के मॉमेंट-जनरेटिंग फलन का प्रयोग करते हुए y_t का आकलित मान y_t के प्रतिबंधित माध्य का मूल्यांकन करके प्राप्त किया जा सकता है



जब प्रक्रिया के पूर्व मान $\{y_t; j \leq k\}$ दिए गए हों:

$$\mu_{y:t|t_k} = \exp(\mu_{Z:t|t_k} + 0.5\sigma_{Z:t|t_k}^2) \quad (8)$$

समीकरण (6) के गॉम्पर्टज़ प्रतिमान के प्राचलों को रूपांतरित प्रक्रिया $\{Z_t\}$ के संयुक्त संभाविता को अधिकतम कर अधिकतम संभाविता (maximum likelihood) पद्धति द्वारा आकलित किया गया। संयुक्त संभाविता प्रतिबंधित संभाविता के काल में के गुणांक है जो कि समीकरण (7) में दिए गए प्रतिबंधित माध्य और परिसरण के साथ गॉशियन हैं। प्रतिमान से आँकड़ों के आकलन के लिए एसएस सॉफ्टवेयर पैकेज में संगणक कोड विकसित किए गए। अंत में, पशु-वार फिट प्रतिमान का मूल्यांकन वर्ग माध्य मूल त्रुटि (आरएमएसई) द्वारा किया गया है:

$$RMSE = [\sum_{t=1}^n (y_t - \hat{y}_t)^2 / n]^{1/2} \quad (9)$$

जहाँ Y_t और \hat{Y}_t क्रमशः t काल में प्रेक्षित और आकलित मान और n काल संख्या को दर्शाते हैं। पशु-वार आरएमएसई का औसत मान आकलित प्रतिमान के प्रदर्शन को दर्शाता है।

परिणाम और चर्चा

उपरोक्त पद्धति का वर्णन करने के लिए आँकड़े भारतीय पशु चिकित्सा अनुसंधान संस्थान, इज्जतनगर, बरेली के शूकर बाड़े से एकत्र किए गए। 210 शूकरों में प्रत्येक का वजन (किलोग्राम) 0, 1, 2, ..., 8 और 12, 16, 20, 24 महीने की उम्र का लिया गया। 20 शूकरों का

यादृच्छिक सैम्पल चुना गया जिसे तालिका-1 में दर्शाया गया है। 20 महीने तक के आँकड़े गॉम्पर्टज़ एसडीई विकसित प्रतिमान के आकलन के लिए और 24वें महीने के आँकड़े आकलित प्रतिमान के प्रदर्शन का मूल्यांकन करने के लिए प्रयोग किए गए। मूल्यांकन करने के लिए वर्ग माध्य मूल त्रुटि (आरएमएसई) की गणना की गई। विकसित किए गए एसएस संगणक प्रोग्राम का प्रयोग कर गॉम्पर्टज़ एसडीई प्रतिमान के प्राचलों का आकलन किया गया तथा आकलित मान तालिका-2 में दर्शाए गए। समीकरण (9) का प्रयोग कर आरएमएसई के औसत मान की गणना 1.35 किलोग्राम की गई जो कि काफी कम है और दर्शाता है कि गॉम्पर्टज़ प्रतिमान का प्रदर्शन विचाराधीन आँकड़ों के लिए उत्तम है। दो चयनित शूकरों के आकलित प्रतिमान का रेखाचित्र प्रेक्षित आँकड़ों के साथ रेखाचित्र-1 में दर्शाया गया है। आकलित प्रतिमान के प्रदर्शन का मूल्यांकन करने के लिए शूकर-वार एक-स्टेप आगे के पूर्वानुमान की गणना की गई और फिर समीकरण (9) का प्रयोग कर शूकर-वार आरएमएसई पूर्वानुमान त्रुटि मान के औसत मान की गणना की गई। यह मान 2.25 किलोग्राम है जो कि काफी कम है और दर्शाता है कि पूर्वानुमान के लिए भी आकलित प्रतिमान का प्रदर्शन अच्छा है।

अतः प्रसंभाव्य विभिन्नात्मक समीकरण पद्धति पशु-वैज्ञानिकों को उनके आँकड़ों के सटीक आकलन में सहायक सिद्ध होगी।

तालिका-1: विभिन्न आयु (महीनों) में शूकरों का वजन (किलोग्राम)

आयु शूकर सं:	1	2	3	4	5	6	7	8	12	16	20	24
1	1.4	2.3	3.4	5.2	6.0	7.8	9.6	10.8	15.0	21.8	30.8	35.0
2	1.3	1.7	2.4	3.0	3.0	3.5	5.1	6.5	10.5	17.0	32.0	37.0
3	1.3	2.1	3.5	4.4	5.4	6.0	6.6	7.2	10.8	13.0	14.0	15.0
4	1.8	2.7	4.0	5.9	6.3	6.8	7.4	8.6	11.8	14.0	13.4	13.4
5	1.7	2.6	4.9	6.3	7.0	6.8	7.4	8.0	11.0	14.0	16.6	18.4
6	1.4	2.2	4.5	5.6	6.8	7.4	8.0	8.6	10.0	13.0	17.0	16.4
7	1.8	2.5	3.4	4.3	4.7	6.0	6.8	6.8	10.4	15.0	17.4	19.4
8	2.1	3.0	4.1	5.2	5.6	7.2	8.2	9.2	10.8	13.8	19.4	20.0
9	1.0	1.6	2.0	2.0	2.2	2.8	4.0	5.0	7.0	11.0	13.8	18.0
10	2.0	3.0	3.6	4.0	4.8	4.4	5.0	6.4	7.4	10.4	16.4	15.6

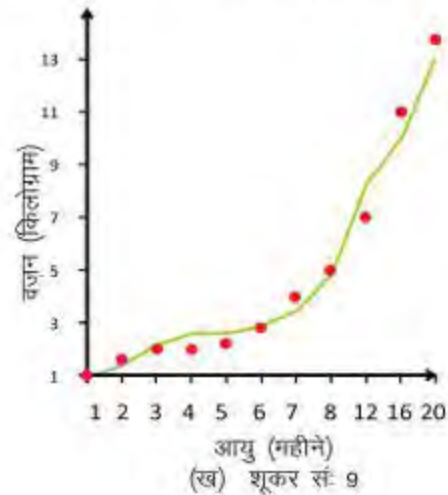
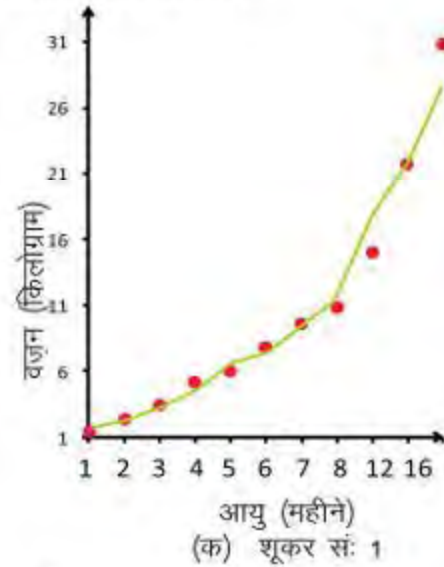


आयु शूकर सं:	1	2	3	4	5	6	7	8	12	16	20	24
11	1.2	2.2	3.0	3.2	4.0	4.2	4.7	5.2	6.4	9.8	15.0	13.6
12	1.7	2.7	3.8	4.7	5.5	7.0	7.8	8.6	13.8	18.6	23.0	23.0
13	2.2	2.6	2.7	3.5	4.6	6.0	7.5	9.8	13.6	20.0	24.0	28.0
14	1.2	2.2	3.0	3.2	4.0	4.2	4.7	5.2	6.4	9.8	15.0	13.6
15	2.1	2.6	3.3	5.4	6.3	7.6	8.6	8.2	11.8	19.2	30.0	35.0
16	1.5	2.2	2.8	3.0	4.3	3.8	4.1	4.8	7.0	15.0	20.4	24.4
17	1.8	2.4	3.7	4.6	6.0	7.2	8.8	9.6	13.0	16.8	21.2	21.0
18	1.6	2.3	3.1	4.3	5.1	6.0	6.6	7.8	12.8	18.6	29.0	33.0
19	2.0	2.8	4.3	5.2	6.1	7.4	8.2	9.4	13.2	16.4	18.4	21.0
20	1.2	2.1	3.0	4.0	4.8	5.6	6.6	7.8	10.8	19.4	27.0	32.0

तालिका-2: आकलित गॉम्पर्टज़ एसडीई प्रतिमान के प्राचलों का आकलन

प्राचल शूकर सं:	α	r	σ^2
1	3.67	0.12	0.008
2	4.03	0.09	0.019
3	2.60	0.23	0.004
4	2.55	0.29	0.008
5	2.63	0.28	0.017
6	2.64	0.27	0.017
7	2.83	0.18	0.008
8	2.72	0.26	0.016
9	2.77	0.13	0.016
10	2.45	0.24	0.032
11	2.34	0.26	0.028
12	3.04	0.20	0.008
13	3.30	0.14	0.020
14	2.34	0.26	0.028
15	3.58	0.12	0.016
16	3.25	0.11	0.024
17	2.99	0.20	0.007
18	3.38	0.15	0.010
19	2.81	0.24	0.009
20	3.42	0.13	0.008

रेखाचित्र-1: शूकर वज़न आँकड़ों के लिए गॉम्पर्टज़ एसडीई प्रतिमान का आकलन





आभार

डॉ. घोष एवं डॉ. प्रज्ञेय अनुसंधान परियोजना सं. एस बी/एस 4/एम एस/ 880/2014 के अंतर्गत उपरोक्त शोध कार्य करने के लिए वित्तीय सहायता प्रदान करने हेतु विज्ञान और इंजीनियरी अनुसंधान बोर्ड (एसईआरबी), नई दिल्ली के आभारी हैं। लेखकगण टंकण हेतु श्रीमती सुदेश अरोड़ा, निजी सहायक के आभारी हैं।

संदर्भ

कोहेन, एस. एवं इलियट, आर. जे. (2015): स्टॉकैस्टिक कैलकुलस एंड ऐप्लीकेशन्स, दूसरा एडिशन, बिरहॉजर, स्विट्जरलैंड।

फीलिपे, पी.ए., ब्रौमान्न, सी.ए., ब्राइट्स, एन. एम. एवं रोक्वेते, सी.जे. (2013): प्रैडिक्शन फॉर इन्डीवीजुअल ग्रोथ इन ए रेन्डम एनवायरमेंट। इन "रीसेंट डेवलपमेंट इन मॉडलिंग एंड ऐप्लीकेशन इन स्टैटिस्टिक्स, एडिटर: ओलिवेरा पी. ई., टेन्डो एम. जी., हेनरिकवीज सी. एवं विची एम., 193-201, स्प्रिंगर, बर्लिन"।

घोष, एच., इकबाल, एम.ए. एवं प्रज्ञेय (2011): बूटस्ट्रेप स्टडी ऑफ पैरामीटर एसटीमेट्स फॉर नॉनलीनियर रिचर्ड्स ग्रोथ मॉडल थ्रू जेनेटिक एल्गोरिथ्म। जर्नल ऑफ अप्लाइड स्टैटिस्टिक्स 38: 491-500।

मैटिस, जे. एच., थॉमस, आर.के., वोपके, वी.डी. डब्ल्यू, ऐलजेन्ड्रो, सी.सी., मैटिस, टी.आई. एवं ग्रांट, डब्ल्यू. ई. (2009): पापुलेशन डाइनामिक्स मॉडल्स बेस्ड ऑन क्युमुलेटिव डेन्सिटी डिपेंडेंट फीडबैक: ए लिंक टु दा लॉजिस्टिक ग्रोथ कर्व एंड ए टेस्ट फॉर सिमिट्री यूजिंग एफिड डाटा। इकोलोजिकल मॉडलिंग 220: 1745-1751।

ओर्सेडल, बी. (2003): स्टॉकैस्टिक डिफरेंशियल इक्वेशन्स: एन इन्ट्रोडक्शन विद ऐप्लीकेशन, पाँचवा एडिशन, स्प्रिंगर साइंस एंड बिजनेस मीडिया, बर्लिन।

प्रज्ञेय एवं कंडाला, वी.एम. (2003): मिक्सड-इन्प्लूएंस नॉनलीनियर ग्रोथ मॉडल। जर्नल ऑफ इंडियन सोसाइटी ऑफ एग्रिकल्चरल स्टैटिस्टिक्स 56: 19-24।

प्रज्ञेय एवं रविचन्द्रन, एस. (2003): ए मेथड फॉर फिटिंग ऑफ नॉनलीनियर फॉक्स मॉडल इन फिशरीज। इंडियन जर्नल ऑफ ऐनिमल साइंसेज 73: 329-331।

सैबर, जी.ए.एफ. एवं वाइल्ड, सी.जे. (2003): नॉनलीनियर रिग्रेशन, जॉन वाइले एंड संस, न्यूयॉर्क।

वेणुगोपालन, आर. एवं प्रज्ञेय (1997): वॉन बर्टलैनफी ग्रोथ मॉडल विद आटोकोरिलेटिड ऐरर। इंडियन जर्नल ऑफ फिशरीज 44: 63-67।

स्थिर लाइनों के उत्पादन के लिए उच्च स्तरीय क्रॉसिंग योजनाएँ

मोहम्मद हारून, सिनी वरगीस, सीमा जग्गी, अनिदिता दत्ता एवं अर्पण भौमिक

सार

संकरण तकनीकी का प्रयोग पौधा एवं पशु उत्पादन क्षेत्र में एक लंबे समय से किया जा रहा है। प्रजनकों को एक ऐसे संकर के उत्पादन में रुचि होती है जो स्थिर एवं प्रतिरोधक क्षमता वाला हो। संकर उत्पादन में मुख्यतः निम्न स्तरीय विधि जैसे द्वि-पथ क्रॉस एवं उच्च स्तरीय विधियाँ जैसे त्रि-पथ और चार-पथ क्रॉस का प्रयोग किया जाता है। उच्च स्तरीय विधियों का प्रयोग कर, हम निम्न स्तरीय विधियों की तुलना में, अभिभावक से संबंधित संयोजन क्षमता के प्रभाव के बारे में अधिक जानकारी प्राप्त कर सकते हैं। यहाँ पर त्रि-पथ एवं चार-पथ क्रॉस को मॉडल द्वारा समझाया गया है एवं आंशिक उच्च स्तरीय क्रॉस योजनाओं के महत्व पर प्रकाश डाला गया है। आंशिक त्रि-पथ एवं चार-पथ क्रॉस योजना के एक-एक उदाहरण द्वारा विभाजन की मात्रा को ज्ञात करने की विधि को समझाया गया है।

मूल शब्द: चार-पथ क्रॉस, द्वि-पथ क्रॉस, विभाजन की मात्रा, संयोजन क्षमता, त्रि-पथ क्रॉस

प्रस्तावना

पौधा एवं पशु उत्पादन के क्षेत्रों में संकरण तकनीकी का प्रयोग एक मुख्य घटक के रूप में किया जाता है। संकरण तकनीकी का मुख्य उद्देश्य अभिभावक लाइनों की आनुवंशिक क्षमता में सुधार कर एक बेहतर संतति का विकास करना होता है। इसी प्रकार किसी भी प्रजनन कार्यक्रम में प्रजनक का मुख्य उद्देश्य भी व्यक्तिगत लाइनों की आनुवंशिक क्षमता में उन्नति करना होता है। किसी भी संकर में एक प्रजनक उच्चतर उत्पादन क्षमता के अलावा स्थिरता एवं प्रतिरोधक क्षमता में रुचि रखता है। इन विशेषताओं का सीधा

संबंध संकर उत्पादन में प्रयोग की गयी तकनीकी से होता है। संकर उत्पादन में मुख्यतः प्रयोग की जाने वाली विधियाँ द्वि-पथ क्रॉस, त्रि-पथ क्रॉस एवं चार-पथ क्रॉस हैं। इन विधियों में से द्वि-पथ क्रॉस विधि सरल होने के कारण सबसे अधिक उपयोग में लायी जाती है। अन्य दो विधियाँ, जो की उच्च स्तरीय विधियाँ हैं, प्रयोग में जटिल होने के कारण प्रजनकों द्वारा कम उपयोग की जाती हैं। उच्च स्तरीय विधियाँ कई प्रकार से निम्न स्तरीय विधियों से बेहतर होती हैं। उच्च स्तरीय विधियाँ निम्न स्तरीय विधियों से आनुवांशिक रूप में अधिक व्यवहार्य, स्थिर एवं निष्पादन में एक रूप होती हैं। उच्च स्तरीय क्रॉस विधियों से उत्पन्न संकरों में द्वि-पथ क्रॉस की अपेक्षा, व्यक्तिगत एवं जनसंख्या के दोनो स्तरों पर, प्रबल प्रतिरोधक तंत्र को प्रदर्शित करते हैं। पशु एवं पौधा उत्पादन क्षेत्रों में हमें कई ऐसे उदाहरण मिलते हैं जहाँ पर उच्च स्तरीय क्रॉस विधियाँ निम्न स्तरीय क्रॉस विधियों से बेहतर प्रदर्शन करती हैं। शुमुगथाई एवं श्री निवासन (2012) के शोध कार्य के अनुसार त्रि-पथ क्रॉस विधि मक्का उत्पादन में प्रयोग की जाने वाली सबसे सामान्य वाणिज्यिक तकनीक है। अगर हम मुर्गी उत्पादन क्षेत्र का उदाहरण लें तो हमें ख्वाजा एवं अन्य (2013) शोध कार्य के अनुसार त्रि-पथ क्रॉस संकरित मुर्गियों में द्वि-पथ क्रॉस की तुलना में कम मृत्यु दर पाया गया है एवं मुर्गियों में अंडों का बड़े आकार एवं अधिक वजन के साथ बेहतर अंडों के लक्षण पाये गए हैं। त्रि-पथ एवं चार-पथ क्रॉस शूकरों के भीट के तेज विकास दर, अच्छी खाद्य दक्षता एवं गुणवत्ता के लिए अब तक की सबसे व्यावहारिक एवं स्वीकार्य विधि है। इसके अलावा यह विधियाँ रेशमकीट उत्पादन के मामले में विषमंगता के दोहन के लिए प्रयोग की जाती हैं।



पूर्ण त्रि-पथ क्रॉस संयोजना

लाईनों की संख्या n के आधार पर बनाई गयी, $(l \times j) \times k$ प्रकार के कुल संभावित त्रि-पथ क्रॉसों के समूह को पूर्ण त्रि-पथ क्रॉस संयोजना के नाम से जाना जाता है। दिये गए d लाईनों से बनाई गयी पूर्ण त्रि-पथ क्रॉस संयोजना में कुल क्रॉसों की संख्या N_1 निम्नलिखित सूत्र द्वारा ज्ञात की जा सकती है:

$$N_1 = \frac{n(n-1)(n-2)}{2}$$

यहाँ पर लाईनों की संख्या $n=5$ के लिए पूर्ण त्रि-पथ क्रॉस संयोजना, जिसमें कुल क्रॉसों की संख्या $N_1 = 30$ को दर्शाया गया है:

$(1 \times 2) \times 3$	$(1 \times 3) \times 2$	$(2 \times 3) \times 1$
$(1 \times 2) \times 4$	$(1 \times 4) \times 2$	$(2 \times 4) \times 1$
$(1 \times 2) \times 5$	$(1 \times 5) \times 2$	$(2 \times 5) \times 1$
$(1 \times 3) \times 4$	$(1 \times 4) \times 3$	$(3 \times 4) \times 1$
$(1 \times 3) \times 5$	$(1 \times 5) \times 3$	$(3 \times 5) \times 1$
$(1 \times 4) \times 5$	$(1 \times 5) \times 4$	$(4 \times 5) \times 1$
$(2 \times 3) \times 4$	$(2 \times 4) \times 3$	$(3 \times 4) \times 2$
$(2 \times 3) \times 5$	$(2 \times 5) \times 3$	$(3 \times 5) \times 2$
$(2 \times 4) \times 5$	$(2 \times 5) \times 4$	$(4 \times 5) \times 2$
$(3 \times 4) \times 5$	$(3 \times 5) \times 4$	$(4 \times 5) \times 3$

त्रि-पथ क्रॉस के लिए मॉडल इस प्रकार है:

$$y_{ijkl} = \mu + h_i + h_j + g_k + d_{ijl} + s_{ijk} + s_{jlk} + t_{lyk} + e_{ijkl}$$

जहाँ पर y_{ijkl} त्रि-पथ क्रॉस की प्रतिक्रिया, μ कुल औसत, h_i एवं h_j अर्ध अभिभावकों की सामान्य संयोजन क्षमता को, g_k पूर्ण अभिभावकों की सामान्य संयोजन क्षमता को, s_{ijk} , s_{jlk} एवं t_{lyk} विभिन्न विशिष्ट संयोजन क्षमताओं को एवं e_{ijkl} त्रुटि को दर्शाते हैं, तथा $i, j, k (i \neq j \neq k) = 1, 2, \dots, n$ है।

पूर्ण चार-पथ क्रॉस संयोजना

चार-पथ क्रॉस संकर किसी दो असंबंधित F_2 संकर के पहली पीढ़ी के संतति होते हैं। इसे $(A \times B)$, $(C \times D)$ के रूप में चिन्हित किया जा सकता है, जहाँ पर A, B, C और D चार अभिभावक एवं (A, B) एवं (C, D) दो F_1 हैं।

दिये गए d लाईनों से बनाई गयी पूर्ण त्रि-पथ क्रॉस संयोजना में कुल क्रॉसों की संख्या N_2 निम्नलिखित सूत्र द्वारा ज्ञात की जा सकती है:

$$N_2 = \frac{n(n-1)(n-2)(n-3)}{8}$$

यहाँ पर लाईनों की संख्या $n=5$ के लिए पूर्ण चार-पथ क्रॉस संयोजना, जिसमें कुल क्रॉसों की संख्या $N_2 = 15$ है, को दर्शाया गया है:

$(1 \times 2) \times (3 \times 4)$	$(1 \times 3) \times (4 \times 5)$	$(1 \times 4) \times (2 \times 5)$
$(1 \times 2) \times (3 \times 5)$	$(1 \times 3) \times (2 \times 5)$	$(1 \times 4) \times (3 \times 5)$
$(1 \times 2) \times (4 \times 5)$	$(1 \times 3) \times (2 \times 4)$	$(1 \times 4) \times (2 \times 3)$
$(1 \times 5) \times (3 \times 4)$	$(1 \times 5) \times (2 \times 3)$	$(4 \times 5) \times (2 \times 3)$
$(2 \times 5) \times (3 \times 4)$	$(1 \times 5) \times (2 \times 4)$	$(2 \times 4) \times (3 \times 5)$

चार-पथ क्रॉस के लिए मॉडल इस प्रकार है:

$$y_{ijkl} = \mu + G_{ijkl} + e_{ijkl}$$

जहाँ पर y_{ijkl} चार-पथ क्रॉस की प्रतिक्रिया, μ कुल औसत, G_{ijkl} अभिभावकों की सामान्य संयोजन तथा विभिन्न विशिष्ट संयोजन क्षमताओं को एवं e_{ijkl} त्रुटि को दर्शाते हैं, और $i, j, k, l (i \neq j \neq k \neq l) = 1, 2, \dots, n$ है।

आंशिक उच्च स्तर (त्रि-पथ एवं चार-पथ) क्रॉस संयोजना

अगर हम कुल क्रॉसों की संख्या को ज्ञात करने वाले सूत्रों पर विचार करें तो हमें यह पता चलता है कि लाईनों की संख्या बढ़ने पर कुल क्रॉसों की संख्या में कई गुना वृद्धि हो जाती है। प्रजनकों के लिए यह स्थिति एक समस्या बन जाती है क्योंकि ऐसी स्थिति में प्रयोग के लागत में भी वृद्धि हो जाती है। इस समस्या के समाधान हेतु हम किसी भी पूर्ण उच्च स्तर (त्रि-पथ एवं चार-पथ) क्रॉस संयोजना का एक प्रतिदर्श जिसे हम आंशिक उच्च स्तर (त्रि-पथ एवं चार-पथ) क्रॉस संयोजना के नाम से जानते हैं, का प्रयोग कर सकते हैं। यहाँ पर हम किसी भी उच्च स्तर (त्रि-पथ एवं चार-पथ) क्रॉस संयोजना से अभिभावकों की संख्या कम ना करके कुल क्रॉसों की संख्या को कम करते हैं और इस प्रकार हमें सभी अभिभावक लाइनों से संबंधित संयोजन क्षमता के प्रभाव के बारे में जानकारी प्राप्त होती है। हिंकलमन ने 1965 में आंशिक त्रि-पथ क्रॉस



कि अवधारणा शुरू कि तथा आंशिक त्रि-पथ क्रॉस को परिभाषित किया। किसी भी दो समान दक्षता वाली आंशिक उच्च स्तर (त्रि-पथ एवं चार-पथ) क्रॉस संयोजना कि तुलना कर उनमें से एक को चुनने के लिए हम विभाजन की मात्रा (f) का प्रयोग किया जा सकता है।

विभाजन कि मात्रा को इस प्रकार परिभाषित किया जा सकता है: यदि n लाइनों के लिए, बनाई गयी आंशिक उच्च स्तर (त्रि-पथ एवं चार-पथ) क्रॉस संयोजना में कुल क्रॉसों की संख्या N_p है, एवं पूर्ण

उच्च स्तर (त्रि-पथ एवं चार-पथ) क्रॉस संयोजना में कुल क्रॉसों की संख्या N_c है तब विभाजन की मात्रा (f) नीचे दिये गए सूत्र से निकाली जा सकती है:

$$f = \frac{N_p}{N_c}$$

उदाहरण 1: यहाँ पर लाइनों की संख्या, $n=7$ के लिए एक आंशिक त्रि-पथ क्रॉस संयोजना (हारुन एवं अन्य, 2016) को प्रस्तुत कर विभाजन की मात्रा को दर्शाया गया है।

(1×2)×3	(1×3)×2	(2×3)×1	(1×3)×5	(1×5)×3	(3×5)×1	(1×4)×7	(1×7)×4	(4×7)×1
(2×3)×4	(2×4)×3	(3×4)×2	(2×4)×6	(2×6)×4	(4×6)×2	(2×5)×1	(2×1)×5	(5×1)×2
(3×4)×5	(3×5)×4	(4×5)×3	(3×5)×7	(3×7)×5	(5×7)×3	(3×6)×2	(3×2)×6	(6×2)×3
(4×5)×6	(4×6)×5	(5×6)×4	(4×6)×1	(4×1)×6	(6×1)×4	(4×7)×3	(4×3)×7	(7×3)×4
(5×6)×7	(5×7)×6	(6×7)×5	(5×7)×2	(5×2)×7	(7×2)×5	(5×1)×4	(5×4)×1	(1×4)×5
(6×7)×1	(6×1)×7	(7×1)×6	(6×1)×3	(6×3)×1	(1×3)×6	(6×2)×5	(6×5)×2	(2×5)×6
(7×1)×2	(7×2)×1	(1×2)×7	(7×2)×4	(7×4)×2	(2×4)×7	(7×3)×6	(7×6)×3	(3×6)×7

बनाई गयी आंशिक त्रि-पथ क्रॉस संयोजना में कुल क्रॉसों की संख्या $N_p=63$ है, पूर्ण त्रि-पथ क्रॉस संयोजना में कुल क्रॉसों की संख्या $N_c=105$ है तथा विभाजन की मात्रा $f=63/105 (=3/5)$ है

उदाहरण 2: यहाँ पर लाइनों की संख्या, $n=8$ के लिए एक आंशिक चार-पथ क्रॉस संयोजना (प्रसाद एवं अन्य, 2005), जिसकी निर्माण विधि संतुलित अपूर्ण खंड अभिकल्पना (8, 14, 7, 4, 3) पर आधारित है, को प्रस्तुत कर विभाजन की मात्रा को दर्शाया गया है।

(2×4) × (6×8)	(1×3) × (5×7)
(3×5) × (7×8)	(2×4) × (6×1)
(4×5) × (1×8)	(3×5) × (7×2)
(5×7) × (2×8)	(4×6) × (1×3)
(6×1) × (3×8)	(5×7) × (2×4)
(7×2) × (4×8)	(6×1) × (3×5)
(1×3) × (5×8)	(7×2) × (4×6)

बनाई गयी आंशिक त्रि-पथ क्रॉस संयोजना में कुल क्रॉसों की संख्या $N_p=14$ है, पूर्ण त्रि-पथ क्रॉस संयोजना में कुल क्रॉसों की संख्या $N_c=210$ है तथा विभाजन की मात्रा $f=14/210 (=1/15)$ है।

साहित्य में अमाज्य एवं अमाज्य घात लाइनों की संख्या के लिए आंशिक त्रि-पथ क्रॉस (हारुन और अन्य, 2016) दास एवं गुप्ता, 1997) एवं आंशिक

चार-पथ क्रॉस (राव्लिंग्स एवं कोकेरहम, 1962; प्रसाद एवं अन्य, 2005) की निर्माण एवं विश्लेषण की विधियाँ उपलब्ध हैं।

निष्कर्ष

यहाँ पर संकर उत्पादन के लिए उच्च स्तरीय क्रॉसिंग योजनाओं का महत्व समझाया गया है। उच्च स्तरीय क्रॉसिंग योजनाओं द्वारा प्रजनक अपेक्षाकृत अधिक स्थिर एवं प्रतिरोधक क्षमता वाली लाइनों का उत्पादन कर सकते हैं। उच्च स्तरीय क्रॉसिंग विधियाँ, जैसे की त्रि-पथ एवं चार-पथ क्रॉस को इस क्षेत्र में उत्तम पाया गया है। साहित्य में आंशिक उच्च स्तरीय क्रॉसिंग योजनाओं की निर्माण विधि एवं विश्लेषण की उपलब्ध विधियों का प्रयोग करके प्रजनक अपने उद्देश्यों को पूरा कर सकते हैं। प्रजनक कम विभाजन की मात्रा वाली आंशिक उच्च स्तर (त्रि-पथ एवं चार-पथ) क्रॉस संयोजना का चुनाव करके कम से कम क्रॉसों की संख्या में ज्यादा लाइनों से संबंधित संयोजन क्षमता के प्रभाव के बारे में जानकारी प्राप्त कर सकते हैं।



संदर्भ

राब्लिंग्स, जे.ओ. एवं कोकेरहम, सी.सी. (1962). डबलक्रॉस संकरितजन संख्या का विश्लेषण. बायो मेट्रिका, 18, 229-244.

हिकेलमेन, के. (1965). आंशिक त्रि-पथ क्रॉसेस. संख्या, 27, 173-196.

पोन्नुरवामी, के. एन. एवं श्री निवास न. एम. आर. (1991). संतुलित अपूर्ण खंड अभिकल्पनाओं की एक श्रेणी का प्रयोग कर आंशिक त्रि-पथ क्रॉसेस का निर्माण. कम्युनिकेशन्स इन स्टैटिस्टिक्स-थियरि एवं मैथड्स, A 20, 3315-3323.

प्रसाद, आर., गुप्ता, वी.के. एवं गुप्ता, आर. (2005). दो-लाईन एवं चार-लाईन प्रयोगों के

लिए इष्टतम अभिकल्पनाएँ. यूटिलिटस मैथमेटिका, 68, 11-32.

ख्वाजा, टी., खान, एस.एच., मुख्तार, एन., परवीन, ए. एवं फरीद, जी. (2013). उप-उष्णकटिबंधीय जलवायु में पारस्परिक F1 संकरित मुर्गियों सहित त्रि-पथ संकरित मुर्गियों की उत्पादन प्रदर्शन, अंडों की गुणवत्ता एवं जैवरसायनिक मापदण्ड. इटालियन जर्नल ऑफ एनिमल साइंसेज, 12 (1), 178-189.

हारून, एम., वर्गीस, सी., वर्गीस, इ. एवं जग्गी, एस. (2016). पशु प्रजनन प्रयोगों के लिए त्रि-पथ क्रॉस अभिकल्पनाएँ. दि इंडियन जर्नल ऑफ एनिमल साइंसेज, 86(6), 94-98.

मेटाजीनोमिक्स: परिचय और अनुप्रयोग

अनु शर्मा, शशि भूषण लाल, द्विजेश चन्द्र मिश्र, नीरज बुढ़लाकोटी, कृष्ण कुमार चतुर्वेदी,
मोहम्मद समीर फारूकी एवं संजीव कुमार

सारांश

मेटाजीनोमिक्स माइक्रोबियल समुदायों की संरचनात्मक और कार्यात्मक रूपरेखा के लिए एक आशाजनक तकनीक के रूप में उभर रहा है। नई अनुक्रमण तकनीकों के आगमन के साथ, कई पर्यावरणीय और मानव माइक्रोबियल समुदायों के अनुक्रमण किये जा चुके हैं जिनको इससे पहले प्रयोगशाला में कल्चर करना मुश्किल था। मेटाजीनोमिक्स तकनीक का परिचय देना और इसके कार्य क्षेत्रों से अवगत करवाना इस लेख का मुख्य उद्देश्य है। आशा है कि यह लेख शोधकर्ताओं और पाठकों को इस तकनीक का एक संक्षिप्त अवलोकन देने में कारगर साबित होगा।

मुख्य शब्द: मेटाजीनोमिक्स, टैक्सोनोमिक विश्लेषण, कार्यात्मक विश्लेषण, तुलनात्मक विश्लेषण

1. परिचय

जीवमंडल में सूक्ष्मजीवों (क्विकमैन आदि, 1998) का वर्चस्व है जो कि मेडिसिन, इंजीनियरिंग और कृषि में प्रायोगिक महत्व रखते हैं (स्लोअन आदि, 2006)। उनके इसी महत्व के कारण, सूक्ष्मजीवों की आनुवांशिक और जैविक विविधता वैज्ञानिक अनुसंधान का एक महत्वपूर्ण क्षेत्र है। सूक्ष्मजीवों के महत्व की स्पष्टता के बावजूद, उनकी विविधता के बारे में बहुत कम जानकारी है। उदाहरण के लिए – पर्यावरण में कितनी प्रजातियां मौजूद हैं, अथवा प्रत्येक प्रजाति क्या करती है और इसके पारिस्थितिक कार्यों की जानकारी (सिंह आदि, 2008)। अभी तक, सूक्ष्मजीवों के संवर्धन में हुई सीमाओं के कारण इन महत्वपूर्ण सवालों के जवाब देने के लिए कोई उपयुक्त तकनीक उपलब्ध नहीं थी। पारंपरिक तरीकों से उन्हीं सूक्ष्मजीवों का

संवर्धन किया जा सकता है जो प्रायोगिक स्थितियों में बढ़ते हैं। हालांकि, यह व्यापक रूप से स्वीकार किया जाता है कि पर्यावरण में 99 प्रतिशत तक सूक्ष्मजीवों को आसानी से कल्चर नहीं किया जा सकता।

इस प्रकार, जैव प्रौद्योगिकी के लिए अधिकांश रोगाणुओं का वर्णन और मूल्यांकन नहीं किया गया है। इस समस्या के समाधान के लिए अलग प्रकार की डी.एन.ए. आधारित आणविक तरीकों को विकसित किया गया है। इन विधियों ने माइक्रोबियल विविधता और पारिस्थितिकी की हमारी समझ को काफी प्रभावित किया है (डीलांग और कार्ल, 2005)। सामान्य तौर पर, 16 एस.आर.आर.एन.ए. (16 S rRNA) जीन विश्लेषण पर आधारित विधियां ने पर्यावरण और वातावरण में मौजूद प्रजातियों के बारे में विस्तृत जानकारी प्रदान करते हैं (स्टीट और स्मिज़, 2004)। 1990 के दशक के बाद, इन कठिनाइयों पर काबू पाने के लिए एक नई तकनीक शुरू की गई है जिसे 'मेटाजीनोमिक्स' कहा जाता है।

“मेटाजीनोम” शब्द का इस्तेमाल पहली बार जे. हैंडेल्समैन द्वारा 1998 में किया गया था जो एक पर्यावरण नमूने में आनुवांशिक सामग्री की कुल राशि के वर्णन के लिए था। मेटाजीनोमिक्स आणविक जीव विज्ञान और आनुवांशिकी को इस प्रकार जोड़ती है जिससे कि पर्यावरण में उपलब्ध नमूनों से आनुवांशिक सामग्री को पहचाना जा सके और उनका वर्णन किया जा सके (चित्र 1)। मेटाजीनोमिक्स एक ऐसा उभरता हुआ क्षेत्र है जिसमें सूक्ष्मजीवों के पूरे समुदाय पर जीनोमिक विश्लेषण कर सूक्ष्मजीव प्रजातियों को प्रयोगशाला में कल्चर किये बिना अलग-अलग किया जा सकता है (पेट्रिक आदि, 2005)।



चित्र 1: मेटाजीनोमिक्स का अवलोकन

2. मेटाजीनोमिक्स में बुनियादी संगणनात्मक कार्य

मेटाजीनोमिक्स में तीन प्रकार के बुनियादी कार्य संगणनात्मक विधियों के द्वारा किए जाते हैं—

- टैक्सोनोमिक विश्लेषण (कौन वहाँ है?)
- कार्यात्मक विश्लेषण (वे क्या कर रहे हैं?)
- तुलनात्मक विश्लेषण (वे कैसे भिन्न हैं?)

2.1 टैक्सोनोमिक विश्लेषण

जैविक अनुक्रमों की उत्पत्ति का वर्णन करने के लिए सामान्य रूप से एन.सी.बी.आई. (NCBI) द्वारा दिया गया वर्गीकरण प्रयोग होता है। यह एक वृक्ष संरचना (Tree Data Structure) पर आधारित एक फाइलोजेनेटिक (Phylogenetic) वर्गीकरण है जो कि जीवों के बीच विकासवादी संबंधों का अनुमान लगाता है (बेन्सन आदि, 2009)। मेटाजीनोमिक्स अनुक्रमों के वर्गीकरण का मुख्य उद्देश्य सटीकता पूर्वक मूल प्रजातियों को खोजना है। हालांकि, यह सामान्यतः संभव नहीं है, विशेष रूप से यदि इस प्रजाति को पहले अनुक्रमित नहीं किया गया है। इसके अलावा, मूल रूप से इन प्रजातियों से संबंधित प्रजाति प्रायः संदर्भ के रूप में उपलब्ध नहीं होती है। ऐसे मामले में, एक टैक्सोनोमिक वर्गीकरण को मूल प्रजातियों के निम्नतम ज्ञात पूर्वज के मुताबिक उच्च वर्गीय रैंक पर पूर्वसूचित (Prediction) करना चाहिए। सिद्धांत रूप में, मेटाजीनोमिक्स शॉटगन अनुक्रमों के टैक्सोनोमिक वर्गीकरण के लिए दो प्रकार के तरीकों को वर्णित किया जा सकता है। संयोजन पर आधारित विधियाँ

पहले अनुक्रम विशेषताओं को निकालती हैं और फिर इनकी तुलना करती हैं, जबकि तुलना-आधारित पद्धति अनुक्रम स्तर पर सीधे संदर्भ अनुक्रमों के साथ मेटाजीनोमिक्स अनुक्रमों की तुलना करती हैं। रचना आधारित टूल्स में से कुछ इस प्रकार हैं—टेट्रा (TETRA), फाइलोजेनेटिक (Phylo Pythi), ताकोआ (TACOA), एन.बी.सी. (NBC) इत्यादि। तुलनात्मक तरीके के आधार पर उपलब्ध टूल्स इस प्रकार हैं—एम.जी.रास्ट (MG-RAST), मेगान (MEGAN), कारमा (CARMA) इत्यादि हैं।

2.2 कार्यात्मक विश्लेषण

मेटाजीनोमिक्स अध्ययनों में कार्यात्मक विश्लेषण मुख्य रूप से दो व्यापक उद्देश्यों के साथ किया जाता है। वे इस प्रकार हैं: सबसे पहले यह निर्धारित करना है कि विभिन्न समुदाय के सदस्यों के कार्यात्मक और मेटाबॉलिक रिपोर्टओयर (Metabolic Repertoires) कार्यक्रम क्या हैं जो उन्हें अलग-अलग प्रभाव डालने में सक्षम बनाते हैं। दूसरा विभिन्न समुदायों के कार्यात्मक रचनाओं में बदलावों की पहचान करना, यदि कोई है। उदाहरण के लिए—वे सूक्ष्मजीवों के समुदाय जो स्वस्थ और बीमार व्यक्तियों में पाए जाते हैं और किसी रोग के कारण संबंधित हो सकते हैं। एक माइक्रोबायोम (Microbiome) की सदस्य प्रजातियों की कार्यात्मक सामग्री निर्धारित करने के लिए सभी (या कम से कम प्रभावी) सदस्यों के लिए कोडिंग और कार्यात्मक क्षमता का व्यापक विश्लेषण किया जाना चाहिए। वैकल्पिक रूप से, यदि अध्ययन का लक्ष्य विभिन्न समुदायों के कार्यात्मक और मेटाबॉलिक रिपोर्टओयर (Metabolic Repertoires) क्षमताओं का विश्लेषण और विषमता दिखाना है, तो समुदायों के लिए कार्यात्मक और मेटाबॉलिक रिपोर्टओयर (Metabolic Repertoires) मार्ग प्रोफाइल तैयार एवं उनकी तुलना की जानी चाहिए। उपयोग किए जाने वाले कुछ महत्वपूर्ण टूल्स हैं—एम जी रास्ट (MG-RAST), मेगान (MEGAN), कारमा (CARMA) इत्यादि।



2.3 तुलनात्मक विश्लेषण

यह कार्य मेटाजीनोम के कार्यात्मक और टैक्सोनोमिक तुलना से संबंधित है। उपयोग किए जाने वाले कुछ महत्वपूर्ण टूल्स हैं—एम जी रास्ट (MG-RAST), मेगान (MEGAN) इत्यादि।

3. मेटाजीनोमिक्स के अनुप्रयोग

मेटाजीनोमिक्स का निम्नलिखित कार्य क्षेत्रों में प्रयोग होता है।

- कोशिकाओं की संरचना और कार्य को समझना
- उपापचय की समझ
- जीनोम इंजीनियरिंग
- दवा और टीका (Vaccine) विकास
- प्रोटीन-प्रोटीन के मध्य परस्पर क्रिया को समझना
- न्यूनतम जीन सेट को परिभाषित करना
- आर.एन.ए./प्रोटीन अभिव्यक्ति की समझ
- डी.एन.ए. की विविधता की खोजना, जीनोटाइपिंग, फॉरेंसिक

4. निष्कर्ष

कई सूक्ष्मजीव अपशिष्ट पदार्थों को विघटित करने, चिकित्सा अनुप्रयोगों के लिए नई दवाएं बनाने, पर्यावरण के अनुकूल प्लास्टिक बनाने या यहां तक कि हमारे खाने योग्य भोजन भी बनाते हैं। इन जीवों से डी. एन.ए. को अलग करके, हम इन प्रक्रियाओं को सुधार सकते हैं और सामाजिक प्रयोग के अनुकूल बना सकते हैं। मेटाजीनोमिक्स का एक अन्य मूल्यवान अनुप्रयोग यह है कि यह प्रभावी ढंग से पानी, मिट्टी और रुमन स्रोत नमूनों की आनुवंशिक विविधता को चिन्हित करने की क्षमता प्रदान करता है। मेटाजीनोमिक्स एक नया और रोमांचक क्षेत्र है जो भविष्य में जैविक विविधता को समझने के लिए एक मानक तकनीक के रूप में विकसित हो सकता है।

संदर्भ

1. क्विटमैन, डब्ल्यू. बी., ए. कोलमन, डी.सी., एवं विबे, डब्ल्यू.जे. (1998). Prokaryotes: जीम unseen

majority. Proc Natl Acad Sci USA, खंड 9 पृष्ठ संख्या 6578-6583.

2. स्लोन, डब्ल्यू. टी., लून, एम., वुडकॉक. एस., हेड, आई. एम., नी, एस. एवं कर्टिस, टी. पी. (2006). Quantifying the role of immigration and chance in shaping prokaryote community structure, पर्यावरण माइक्रोबायोलॉजी, खंड 8, पृष्ठ संख्या 732-740.
3. सिंह, बी., गौतम, एस. के., वर्मा, वी., कुमार, एम. एवं सिंह बी., (2008). Metagenomics in animal gastrointestinal ecosystem: Potential biotechnological prospects, Anaerobe (एनारोब.), खंड 14 पृष्ठ संख्या 138
4. हैंडेल्समैन, जे., रांडन, एम. आर., ब्रैडी, एस. एफ., क्लैडी, जे. एवं गुडमैन, आर. एम. (1998), Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products, न रसायन विज्ञान और जीवविज्ञान, खंड 5, संख्या 10, पृष्ठ संख्या R245-R249
5. डीलांग, ई. एफ. एवं कार्ल, डी. एम. (2005). Genomic perspectives in microbial oceanography. Nature न (प्रकृति), खंड 437, पृष्ठ संख्या 336-342.
6. स्ट्रेइट, डब्ल्यू. आर., एवं कार्ल, डी. एम. (2004). Metagenomics - the key to the uncultured microbes. Curr. Opin. Microbiol., एम खंड 7, पृष्ठ संख्या 492-498
7. पैट्रिक डी, श्रलॉस, एवं हैंडेल्समैन, जे. (2005). Metagenomics for studying unculturable microorganisms: cutting the Gordian knot, Genome Biol. ts- (जीनोम बीओएल), खंड 6, पृष्ठ संख्या 229
8. बेन्सन, डी. ए., कर्शमिज्जाजी - आई., लिपमानडी, जे., ओस्टेल, जे., एवं सैयर्स, ई. डब्ल्यू. (2009) GenBank ए., न्यूक्लिक एसिड रिस, खंड 37 डाटाबेस इशु: (पृष्ठ संख्या डी26-डी31).

भारतीय कृषि अनुसंधान परिषद (आईसीएआर) में एंटरप्राइज रिसोर्स प्लानिंग सिस्टम (आईसीएआर-ईआरपी) का कार्यान्वयन

मुकेश कुमार, सुदीप मारवाह, अलका अरोड़ा,
अंशु भारद्वाज, अंजनी कुमार चौबे एवं शशि दहिया

ईआरपी-अवलोकन

किसी भी उद्यम प्रणाली की तरह ईआरपी, संगठन की व्यावसायिक प्रक्रियाओं को स्वचालित करता है और किसी भी संगठन में ई-शासन का एक आवश्यक घटक बन जाता है। ईआरपी आम तौर पर व्यापक, जटिल और अनुकूलन योग्य अनुप्रयोग है जो मुख्य व्यवसाय प्रक्रियाओं और संगठन के मुख्य प्रशासनिक क्षेत्रों का समर्थन करते हैं। ईआरपी कार्यान्वयन संगठन की आवश्यकता के अनुसार अनुकूलन को संदर्भित करता है और यह अपने घटक के सभी चरणों में से सबसे अधिक कठिन है। इस चरण में नियमित व्यवसाय प्रभावित हो जाता है। कर्मचारियों को काम करने के अपने स्थापित तरीकों पर अत्यधिक बोझ और चुनौती बन जाती है। प्रभावी क्रियान्वयन के लिए व्यावसायिक प्रक्रियाओं को फिर से रिइंजीनियरिंग/पुनः संरचना की आवश्यकता है।

आईसीएआर में ईआरपी के कार्यान्वयन के लिए सामरिक निर्णय

भारतीय कृषि अनुसंधान परिषद (आईसीएआर), जिसमें 110 केंद्रीय संस्थानों/राष्ट्रीय अनुसंधान केंद्र और उनके क्षेत्रीय स्टेशन शामिल हैं, जो सरकारी नियमों और विनियमों का पालन करता है। आईसीएआर स्तर पर संस्थानों और समेकन द्वारा रिपोर्टिंग के मैन्युअल सिस्टम की अवरोधों का एहसास हो गया। आईसीएआर अधिक प्रभावी तरीके से बजटीय नियंत्रण पर निगरानी रखना और स्पष्ट, सटीक और समय पर रिपोर्टिंग के साथ पूरक करना चाहता है।

शुरूआत में, वित्तीय प्रबंधन प्रणाली को एनआईएपी परियोजनाओं के लिए 7 पायलट संस्थानों में परामर्शदाता

की मदद से लागू किया गया था।

पिछले अनुभव के आधार पर: समाधान

आईसीएआर ने ओरेकल आर 12.1.3 के अनुकूलन के माध्यम से प्रक्रियाओं, मानव संसाधन, परियोजनाओं, सामग्री और वित्त का प्रबंधन करने के तरीके को बदलने के लिए एक रणनीतिक निर्णय लिया। तदनुसार, आईसीएआर की जरूरतों के अनुसार अनुकूलित आईसीएआर-ईआरपी समाधान मुख्यालय सहित सभी संस्थानों में लागू किया गया है।

आईसीएआर-ईआरपी समाधान का कार्यान्वयन कार्यप्रणाली

आईएएसआरआई को 6 संस्थानों के कंसोर्टियम के प्रमुख संस्थान के रूप में जनवरी 2012 को आईसीएआर-ईआरपी कार्यान्वयन के लिए आईबीएम के सलाहकार की सहायता से जिम्मेदारी सौंपी गयी थी। समाधान के अनुकूलन के मार्गदर्शन और निगरानी के लिए कार्यात्मक 5 कोर समूह का गठन किया गया। अनुकूलन के दौरान, प्रयोक्ता स्वीकृति परीक्षण, प्रशिक्षण और कार्यान्वयन प्रक्रिया, विभिन्न दस्तावेज (जैसी है जैसे-प्रक्रिया, होने वाली प्रक्रिया, तैनाती वास्तुकला, नियमावली, कंप्यूटर आधारित प्रशिक्षण सामग्री और लघु ई-मैन्युअल) तैयार किए गए हैं। आईसीएआर-ईआरपी की होस्टिंग के लिए आईएएसआरआई डाटा सेंटर में सर्वर और सैनस्थापित किए गए हैं।

ओरेकल ईआरपी

एंटरप्राइज रिसोर्स प्लानिंग (ईआरपी) प्रणाली संपूर्ण संगठन में जानकारी के आंतरिक और बाहरी प्रबंधन को एकीकृत करती है जिसमें बजटिंग, वित्तीय प्रबंधन,



परियोजना प्रबंधन, प्रोक्योरमेंट और स्टोर प्रबंधन, मानव संसाधन प्रबंधन और पेट्रोल शामिल हैं।

मुख्य विशेषताएं

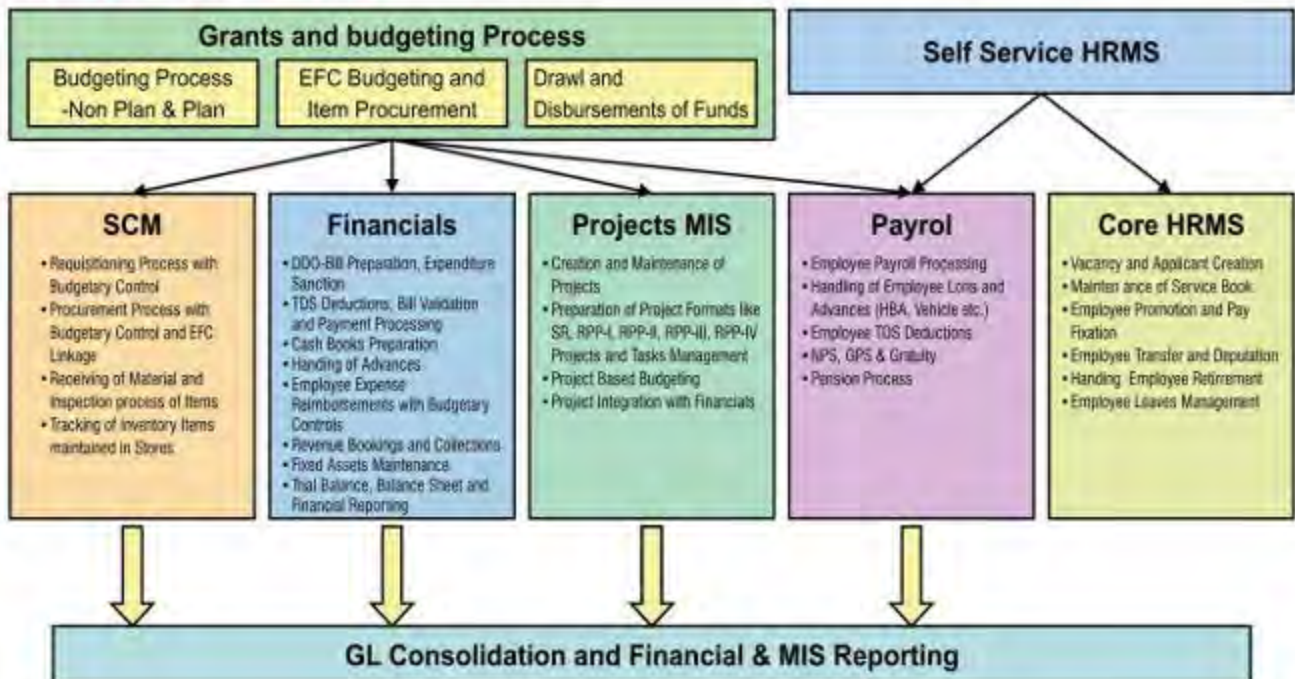
- **एकीकरण और स्वचालन** : ईआरपी सिस्टम संगठन की सीमाओं के अंदर सभी व्यवसाय कार्यों (जैसे वित्त, प्रशासन, वैज्ञानिक, तकनीकी) के बीच सूचना के प्रवाह को सुगम बनाता है और बाहरी हिस्सेदारों (जैसे आपूर्तिकर्ता, बैंक आदि) के लिए कनेक्शन का प्रबंधन करता है।
- **उत्तम व्यवहार और नियंत्रण** : प्रक्रिया अखंडता खोने के बिना प्रत्येक व्यावसायिक प्रक्रिया को पूरा करने के लिए सबसे प्रभावी तरीके से एक मंच प्रदान करता है।
- **रिपोर्टिंग** : उद्योग मानक के अनुसार सटीक और समान रिपोर्टिंग।
- प्रत्येक मॉड्यूल में एक सुसंगत रूप और अनुभव प्रदान करता है।
- **स्केलेबल** : नए मॉड्यूल और फंक्शनलिटीज़ को निरंतर आधार पर जोड़ा जा सकता है।

आईसीएआर-ईआरपी: उद्देश्य

निम्नलिखित व्यापार प्रक्रिया के क्षेत्रों में एंटरप्राइज संसाधन समाधान को लागू करने के लिए योजना।

- **वित्तीय प्रबंधन** : सामान्य खाताधारक, लेखा भुगतान योग्य, लेखा प्राप्य, नकद प्रबंधन, फिक्स्ड आस्तियों प्रबंधन, बजट प्रबंधन और अनुदान।
- **परियोजना प्रबंधन** : परियोजना की जानकारी, लागत, परियोजना दस्तावेज, अनुबंध प्रबंधन और परियोजना दस्तावेजों के सहयोग।
- **आपूर्ति श्रृंखला प्रबंधन** : खरीद और इन्वेंट्री प्रबंधन।
- **मानव संसाधन प्रबंधन** : कर्मचारी की जानकारी, मानव संसाधन नीतियों, प्रबंधन, निष्पादन और मूल्यांकन प्रणाली।
- **पेट्रोल सिस्टम** : वेतन, जीपीएफ, पेंशन भुगतान, सेवानिवृत्ति लाभ गणना और आयकर गणना।
- **स्व सेवा मानव संसाधन** : कार्मिक और व्यावसायिक सूचना, जीपीएफ, छुट्टी, वेतन आदि।

आईसीएआर-ईआरपी: स्कोप



ए) एचआरएम और पेरोल मॉड्यूल एचआर प्रक्रियाएं: अवलोकन



कोर एचआरएमएस फॉर्मस कर्मचारी बुनियादी विवरण

- नाम, आयु, वैवाहिक स्थिति, पैन, संपर्क नंबर, ईमेल आदि।
- स्थायी पता, पत्राचार, होम टाउन पता।
- व्यक्तिगत सूचना – धर्म, पासपोर्ट सूचना आदि।
- शैक्षिक योग्यता विवरण, आश्रित और नामांकित विवरण।
- चिकित्सा आकलन इतिहास विवरण, पुराना कर्मचारी आईडी, सर्विस बुक नंबर, कार्मिक संख्या।

कर्मचारी सेवा विवरण

- परिवीक्षा, पदोन्नति, स्थानांतरण आदि, वेतन विवरण, वेतन परिवर्तन, पर्यवेक्षक विवरण आदि, अधिवक्ता, स्वैच्छिक सेवानिवृत्ति आदि।
- लीव इतिहास विवरण और उनकी सेवा के दौरान सभी परिवर्तनों सहित कई कर्मचारी संबंधित जानकारी।

स्वयं सेवा लेनदेन

- पता अपडेट, वैवाहिक स्थिति और अंतिम नाम।
- विगत और वर्तमान सेवा विवरण देखना।
- शिक्षा योग्यता का विवरण अद्यतन करें।
- पत्ते को लागू करना और रिपोर्ट में शामिल होना सबमिट करना।
- होम टाउन बदलें
- प्रशिक्षण, प्रतिनियुक्ति और अन्य गतिविधियों को अद्यतन करना
- पासपोर्ट, विदेशी यात्रा, उच्च शिक्षा के लिए एनओसी
- उच्चतर अध्ययन, अन्य रोजगार आदि के लिए आवेदन
- वार्षिक कर घोषणा
- संपत्ति घोषणा
- नामांकित घोषणा
- वेतन पर्ची देखना



आईसीएआर-ईआरपी: होम पेज



आईसीएआर एचआर रिपोर्ट (आईसीएआर ईआरपी)

प्रबंधन स्तर

- आईसीएआर एचआर वेतन पची रिपोर्ट
- आईसीएआर एचआर छुट्टी/लीव रिपोर्ट
- आईसीएआर मानव संसाधन हस्तांतरित कर्मचारी रिपोर्ट
- आईसीएआर मानव संसाधन कार्यक्रम में शामिल प्रशिक्षण प्रशिक्षण रिपोर्ट
- आईसीएआर कर्मचारी विस्तार रिपोर्ट
- आईसीएआर एचआर वैज्ञानिक मंजूर शक्ति रिपोर्ट
- आईसीएआर एपीआर कर्मचारी रेटिंग रिपोर्ट
- आईसीएआर मानव संसाधन जाति आधारित हेड काउंटी रिपोर्ट
- आईसीएआर मानव संसाधन डिसिपीलीन वाईज रिपोर्ट
- आईसीएआर एचआर प्रोबेशन स्टेटस रिपोर्ट
- आईसीएआर एचआर धर्म वार प्रमुख गणना रिपोर्ट
- आईसीएआर एचआर संवर्धन रिपोर्ट
- सेवा रिपोर्ट के दौरान आईसीएआर एचआर लीव एनकैशमेंट

उपयोगकर्ता स्तर

- आईसीएआर एचआर वेतन पची रिपोर्ट
- आईसीएआर एचआर कर्मचारी प्रोफाइल रिपोर्ट

- आईसीएआर एचआर छुट्टी कार्यालय आदेश
- आईसीएआर एचआर योग्यता अवधि सेवा
- आईसीएआर एचआर कार्यालय ऑर्डर होम बदलें
- आईसीएआर एपीआर कर्मचारी रेटिंग रिपोर्ट
- आईसीएआर एचआर संवर्धन कार्यालय आदेश
- आईसीएआर वार्षिक वृद्धि कार्यक्रम
- आईसीएआर एलटीसी कार्यालय आदेश रिपोर्ट
- आईसीएआर कर्मचारी वेतनलेख डाटा
- आईसीएआर मुख्यालय पेरोल कर्मचारी वाइज पे रजिस्टर रिपोर्ट
- आईसीएआर एचआर ट्रांसफर ऑर्डर
- आईसीएआर एचआर पुष्टिकरण कार्यालय आदेश
- आईसीएआर एचआर कर्मचारी योग्यता रिपोर्ट

एचआरएम के लाभ

एचआरएम का मुख्य लाभ कारगर गति से बेहतर गुणवत्ता सेवाओं और कुशल सेवाएं है।

- नियमित कार्य के लिए अच्छा है जैसे रिकॉर्ड रखने और बनाए रखने के लिए पोर्टफोलियो।
- यह लागत, समय और श्रम के कटौती में मदद करता है।
- सटीकता में सुधार और मानव पूर्वाग्रह को कम करता है।
- व्यापार प्रक्रिया, उत्पादकता और कर्मचारी संतुष्टि में एक और अधिक गतिशील वर्कफ्लो।
- सब कुछ मानकीकृत और स्वचालित है ताकि सभी को लाभ हो सके।
- एचआरएमएस एक कागज रहित कार्यालय की ओर निर्णायक कदम निमाता है।

बी) वित्त और बजट मॉड्यूल

संस्थान स्तर

- बजट की तैयारी (ईएफसी, गैर योजना, योजना)
- संस्थान विशिष्ट बजट रिपोर्ट
- आईसीएआर से धन की प्राप्तियां



- अप्रयुक्त धन के आईसीएआर को प्रेषण
- संस्थान विशिष्ट व्यय रिपोर्ट
- संस्थान के भीतर नामित प्रमुखों में फंड का पुनः अनुमोदन

आईसीएआर मुख्यालय स्तर

- बजट का एकीकरण (ईएफसी, गैर योजना, योजना)
- ड्राफ्ट एसबीई
- बजट कट कार्य
- समेकन और अंतिम एसबीई
- अंतिम बजट के लिए संस्थानों को सूचना (आरई या बीई)
- फंड का ड्रॉल और निधि वितरण करना
- मनोनीत प्रमुखों के अंतर्गत एसएमडी के भीतर संस्थानों में फंड का पुनः प्राप्ति
- संस्थानों में सभी बजट और व्यय रिपोर्टों को देखना।

वित्त समाधान: अवलोकन

खरीदी प्रक्रिया

- बजटीय नियंत्रण के साथ मांगना प्रक्रिया
- बजट नियंत्रण और ईएफसी संबंधन के साथ पीओ निर्माण प्रक्रिया।
- वस्तुओं की सामग्री और निरीक्षण प्रक्रिया प्राप्त करना।
- स्टोरों में रखी सूची आइटमों का ट्रैकिंग।

वित्त-डीडीओ चक्र और लेखा परीक्षा और लेखा

- डीडीओ—बिल तैयारी, व्यय स्वीकृति
- टीडीएस कटौती, विधेयक मान्यकरण और भुगतान प्रसंस्करण
- नकद पुस्तकें तैयारी
- अग्रिमों का संचालन
- बजट नियंत्रण के साथ कर्मचारी व्यय प्रतिपूर्ति
- राजस्व बुकिंग और संग्रह
- स्थिर संपत्ति रखरखाव

- ट्रायल बैलेंस, बैलेंस शीट और समेकित वित्तीय रिपोर्टिंग

वित्त और पेरॉल एकता

- कर्मचारी पेरॉल संसाधन
- पेंशन प्रसंस्करण
- जीपीएफ और गैरच्युटी
- कर्मचारी ऋण और अग्रिम

वित्त और बजट मॉड्यूल के लाभ

- 1) प्रमुख क्षेत्रों में सिस्टम के माध्यम से स्वचालित रिपोर्ट उत्पादन, उनमें से कुछ नीचे हैं।
 - विधेयक रिपोर्ट में स्पष्ट रूप से विक्रेता की जानकारी के साथ बनाए गए बिलों का विवरण दिखाया गया है।
 - इलेक्ट्रॉनिक फंड ट्रांसफर रिपोर्ट जो भुगतान करने के लिए उपयोग की जाती है।
 - सिस्टम के जरिए TR-5 रिपोर्ट उत्पन्न हो जाती है जिसे खरीदार को रसीद के रूप में दिया जा सकता है।
 - कैश बुक स्वचालित रूप से उत्पन्न हो जाती है।
- 2) डीडीओ ईआरपी के माध्यम से बिलों को प्रोसेस करने से पहले उपलब्ध बजट की जांच कर सकता है।
- 3) आईसीएआर मुख्यालय सभी संस्थानों के लिए एक केंद्र स्थान से सभी वित्तीय लेनदेन की निगरानी कर सकते हैं।
- 4) सभी संस्थानों के लिए बजट एक केंद्र स्थान से दर्ज किया जा सकता है।

सी) आपूर्ति श्रृंखला प्रबंधन मॉड्यूल

इलेक्ट्रॉनिक सप्लाइ चैन प्रबंधन व्यवसाय-टू बिजनेस (बी 2 बी) और बिजनेस-टू-ग्राहक (सीसी 2) प्रक्रियाओं को बढ़ाने और गति, चपलता, रीयल-टाइम कंट्रोल, और ग्राहक संतुष्टि को बेहतर बनाने के लिए सहयोग करता है। एससीएम मॉड्यूल इंडेंट के निर्माण से सामग्री के मुद्दे से प्रक्रिया की जानकारी के प्रवाह के साथ काम करता है।



एससीएम मॉड्यूल जिसमें सामग्री प्रबंधन प्रक्रिया में सुधार होता है जिसमें दो घटक शामिल होते हैं: खरीद और भंडार।

खरीद में, प्रमुख आपरेशन जैसे:

- परचेज रेक्विजीशन
- मांग निर्माण
- खरीद की मांग डेटा दर्ज करना
- अपेक्षित करों में प्रवेश करना
- संलग्नक रखकर
- अनुरोध की स्वीकृति प्रक्रिया
- खरीद की मांग निर्माण

आरएफक्यू (कोटेशन के लिए अनुरोध)

- प्रक्रिया अवलोकन
- आरएफक्यू बनाना
- आरएफक्यू के लिए आपूर्तिकर्ता जोड़ना

कोटेशन

- प्रक्रिया अवलोकन विवरण
- कोटेशन बनाना
- उद्घरण विश्लेषण और अनुमोदन
- खरीद आदेश

एक प्राप्ति से मानक खरीद आदेश (ऑटो बनाने के माध्यम से)

- पीओ में कर दर्ज करना
- खरीद आदेश को मंजूरी
- पीओ के संशोधन
- एक पीओ रद्द करना
- दर संविदाएं
- दर अनुबंध को मंजूरी
- ब्लैकट रिलीज

प्रदायक निर्माण

- प्रदायक/आपूर्तिकर्ता साइट निर्माण
- क्रय रिपोर्ट चलाना

• फंड उपलब्ध इंडेंट पर चेक सिस्टम संस्थान के भीतर स्टॉक की उपलब्धता, प्राप्त/निरीक्षण/वितरण प्रक्रिया, रिपोर्ट, भौतिक समस्या और भौतिक स्थानांतरण का प्रबंधन करता है।

- संगठन पदानुक्रम
- मद मास्टर
- आइटम वर्गीकरण
- आइटम संहिताकरण
- आइटम सृजन

प्राप्त और लेनदेन प्राप्त करना

- एक खरीद आदेश के खिलाफ प्राप्त करना
- प्राप्त करने के समय में करों का संशोधन
- निरीक्षण
- लोट नियंत्रित आइटम प्राप्त करना
- वितरण
- एलसी विवरण के साथ विदेशी विक्रेताओं से प्राप्त करना
- मानक रसीद की अस्वीकृति

विक्रेता को वापस करें

- वेंडर को स्टैंडर्ड रिटर्न
- अस्वीकृत सामग्री को वापिस करना

विविध लेनदेन

- कर्मचारी को सामग्री जारी
 - परियोजना के लिए सामग्री को जारी करना
- संस्थान स्टोरों का उत्तरदायित्व – सामग्री जारी चल रहा इन्वेंटरी रिपोर्टें इन्वेंट्री अवधि समापन।

आपूर्ति श्रृंखला प्रबंधन रिपोर्ट

- आईसीएआर फंड की पूछताछ रिपोर्ट
- आईसीएआर खरीद आदेश प्रिंट रिपोर्ट
- आईसीएआर खर्च स्वीकृत रिपोर्ट
- आईसीएआर खरीद आदेश विवरण रिपोर्ट
- आईसीएआर सामग्री रिलीज रिपोर्ट
- आईसीएआर रसीद रजिस्टर रिपोर्ट



प्रोक्वोर्मेंट और स्टोर के लाभ

- चालान निर्माण के स्वचालन और खरीद आदेश या रसीद के माध्यम से मिलान।
- सभी सामग्री को ईआरपी में कब्जा कर लिया जाना चाहिए ताकि हाथ में स्पष्ट तस्वीर स्टॉक हो।
- इन्वेंटरी लेनदेन रजिस्टर किसी भी मद के सभी लेनदेन दे रहा है।
- कर्मचारियों को जारी किए गए सामान की आसान ट्रैकिंग।

डी) परियोजना प्रबंधन मॉड्यूल

2012 में आईसीएआर द्वारा जारी नवीनतम दिशानिर्देशों के अनुसार, नए आरपीपी प्रारूपों को नए मूल्यांकन और मूल्यांकन दस्तावेजों के साथ नए आरपीपी प्रारूपों के द्वारा बदल दिया गया है। यह समाधान नवीनतम दिशानिर्देशों पर आधारित है।

परियोजनाएं के लाभ

- बजट वित्तपोषण को ट्रैक करने के लिए लगभग सभी संस्थानों में उपयोग किया जाता है
- बजट बनाम व्यय रिपोर्ट आसानी से परियोजनाओं और संबंधित व्यय पर स्पष्ट डेटा देता है।
- सहयोगात्मक परियोजनाओं के मामले में प्रमुख संस्थान व्यय संस्थान के अनुसार ट्रैक कर सकते हैं।
- सभी परियोजनाओं के लिए व्यय बिल को ट्रैक करने के लिए व्यय रिपोर्ट का विवरण।

आईसीएआर-ईआरपी कार्यान्वयन में प्रमुख चुनौतियां
मानव संसाधन और पेट्रोल मॉड्यूल : बुनियादी अनिवार्य डेटा को सत्यापित किया गया है और सिस्टम में अपडेट किया गया है।

- एचआर लीड मॉड्यूल को स्थापना अनुभाग द्वारा सभी कर्मचारियों की प्रत्येक प्रकार की छुट्टी के लिए रिपोर्टिंग अधिकारी और केंद्र शेष छुट्टी को अद्यतन करना आवश्यक है।

- पेट्रोल मॉड्यूल को लागू करने के लिए मानव संसाधन मॉड्यूल आवश्यक है।
- कार्यान्वित करने के लिए अन्य कार्यविधियों को सेवा पुस्तक से संबंधित डेटा अपडेट की आवश्यकता है। उदाहरण के लिए आईसीएआर द्वारा आवश्यक सभी संस्थानों द्वारा प्रशिक्षण डेटा प्रस्तुत किया गया है।
- इंस्टीट्यूट और बैंक के पेंशनर डेटा को सुलझाने की जरूरत है।

आपूर्ति शृंखला प्रबंधन : इंडेंट सृजन, पीओ, स्टॉक और भौतिक मुद्दे को एक साथ लागू करने की आवश्यकता है। जबकि संस्थान अपने आराम के अनुसार संयोजन को लागू कर रहे हैं।

परियोजना प्रबंधन : आवश्यकता के अनुसार रूप में सभी संस्थानों में परियोजना में बजट प्रबंधन लागू किया गया है। विभिन्न परियोजनाएं बनाई जा रही हैं लेकिन आरपीपी प्रारूप विवरण जैसी अन्य सुविधा अपडेट नहीं की जा रही है।

वित्त मॉड्यूल : सभी कार्यशीलताएं अब पूरी तरह चालू हैं, इसके अलावा वार्षिक खाता तैयार करना, जिसके लिए निम्नलिखित की आवश्यकता होती है

- आईसीएआर को पिछले वर्ष यानी 31 मार्च को प्रस्तुत वार्षिक खाते के साथ एसेट डेटा का समाधान।
- 31 मार्च को अगले वर्ष के लिए वार्षिक खातों के समाधान के लिए प्रणाली द्वारा गणना मूल्यहास के मिलान की सुविधा के लिए खरीद और खाते के प्रमुख (नॉनप्लान/योजना/परियोजना) की तिथि के साथ परिसंपत्ति खातों की तैयारी।
- लेनदेन के खाते प्रमुखों के संबंध के दौरान किए गए लेनदेन में त्रुटियों का सुधार।

सभी लेनदेन को सिस्टम में दर्ज करने की आवश्यकता है और मैनुअल बुक रखने से बचा जाना चाहिए।

- संस्थान की लेखापरीक्षा और वित्त टीम तिमाही समापन के लिए योजना बनाई जाएगी।



- किसी भी मैन्युअल गलती से बचने के लिए ईआरपी के माध्यम से सभी बिल का भूगतान
- फिक्स्ड एसेट कैपिटलाइजेशन ईआरपी के माध्यम से किया जाना चाहिए जो फिक्स्ड एससेट अनुसूची और फिक्स्ड एसेट रजिस्टर की स्थापना में मदद करेगा।

आईसीएआर-ईआरपी कार्यान्वयन में सामान्य बाधाएं

- **आईसीटी इन्फ्रास्ट्रक्चर** : शुरुआती पर्याप्त कंप्यूटिंग और लैन इन्फ्रास्ट्रक्चर और जावा कॉन्फिगरेशन के उपयोगकर्ताओं के स्तर पर प्रशासनिक/वित्त विभागों में जानकारी की कमी।
- **इंटरनेट कनेक्टिविटी की उच्च गति** : 50% संस्थानों (पूर्वोत्तर राज्यों, एनआरसीसी, एनआरसीई आदि) और आंतरिक फायरवॉल मुद्दों (बड़े और मध्यम संस्थान) पर आवश्यक बैंडविड्थ की कमी के साथ-साथ समाधान तक पहुंचने में बार-बार कनेक्टिविटी का लगातार टूटना।
- **परिवर्तन प्रबंधन** : उपयोगकर्ताओं को इतना उपयोगकर्ता के अनुकूल नहीं लगता। संचालन कर्मचारियों में आईटी ज्ञान की कमी के कारण कार्य में धीमी गति, कर्मियों की औसत उम्र उच्च (45-48) है, विश्वसनीयता के लिए मैन्युअल सिस्टम के साथ प्रणाली का उपयोग करने की लंबी अवधि।
- **डिजिटलाइजेशन और डाटा का परीक्षण** : डेटा प्रविष्टि ऑपरेटर्स की कम समझ के कारण डिजिटलीकरण में देरी, मैन्युअल रिकॉर्ड अपडेट करने के कारण संस्थान के कर्मियों द्वारा पूर्ण अभिलेखों की उपलब्धता और सत्यापन समय पर नहीं करना।
- आवेदन की आसान पहुंच के लिए प्रोजेक्ट तैयार करने में खामियां विशेष रूप से समय सारिणी, पर्याप्त उपयोगकर्ता-अंत हार्डवेयर और नेटवर्क के

- बुनियादी ढांचे और संस्थान स्तर पर कॉन्फिगरेशन।
- स्वीकृति प्रक्रिया के कई स्तर और अनुमोदन में प्रक्रियात्मक विलंब और कटौती के कारण लेनदेन की प्रक्रिया बढ़ी।
- उपयोगकर्ता समर्थन को एक महीने से अधिक ऑनसाइट की आवश्यकता होती है।

आईसीएआर-ईआरपी कार्यान्वयन के प्रभाव/लाभ

कार्यान्वयन सेवा उन्मुख है, इसमें सभी आईसीएआर के कर्मचारी उपयोगकर्ता हैं। भविष्य में यह आईसीएआर की जीवन रेखा होगी।

अत्यावधि:

- आईटी सशक्तीकरण।
- बेहतर सूचना प्रबंधन।
- व्यावसायिक प्रक्रियाओं और समान मानकीकरण।
- प्रक्रिया वर्कफ्लो का स्वचालन।
- एकीकरण और डेटाबेस के साझाकरण।
- समान रिपोर्टिंग।
- संसाधनों की बेहतर योजना।
- व्यावसायिक प्रक्रियाओं में पारदर्शिता।
- अंतर-संस्थागत सहयोग।

लंबी अवधि:

- सिस्टम में नियंत्रण और प्रभावीता हासिल करना
- प्रक्रिया में संगठनात्मक दक्षता, उत्पादकता और पारदर्शिता में वृद्धि।
- व्यापार प्रक्रियाओं के पुनः इंजीनियरिंग के लिए अवसर प्रदान।
- प्रबंधन अधिकारियों/निर्णय निर्माताओं के लिए पेपर, निम्न कार्यालय प्रबंधन और डैश बोर्ड की ओर कदम।

सांख्यिकी-विमर्श

2017-18

अंक
13

राजभाषा खण्ड



संस्थान की राजभाषा यात्रा : 2017-18

ऊषा जैन

भा.कृ.अनु.प. – भारतीय कृषि सांख्यिकी अनुसंधान संस्थान में हिन्दी के प्रगामी प्रयोग में महत्वपूर्ण अभिवृद्धि हो रही है। संस्थान द्वारा समस्त प्रशासनिक कार्य शत-प्रतिशत हिन्दी में और यथा आवश्यक द्विभाषी हो रहा है। राजभाषा नीति को संस्थान में सुचारु रूप से कार्यान्वित किया जा रहा है। भारत सरकार, गृह मंत्रालय, राजभाषा विभाग द्वारा जारी वार्षिक कार्यक्रम में निहित लक्ष्यों को संस्थान में लगभग पूरा कर लिया गया है। संस्थान द्वारा समस्त प्रशासनिक कार्य शत-प्रतिशत हिन्दी में और यथा आवश्यक द्विभाषी हो रहा है।

भारत सरकार, राजभाषा विभाग की नगर राजभाषा कार्यान्वयन समिति (उत्तरी दिल्ली) द्वारा संस्थान में हिन्दी के प्रगामी प्रयोग से सम्बन्धित मार्च 2017 एवं सितम्बर 2017 को समाप्त छःमाही रिपोर्टों के आधार पर संस्थान को "उत्कृष्ट श्रेणी" में वर्गीकृत किया गया।

नगर राजभाषा कार्यान्वयन समिति (उत्तरी दिल्ली) द्वारा सदस्य कार्यालयों के लिए आयोजित प्रश्नोत्तरी प्रतियोगिता के प्रथम चरण में संस्थान के प्रतिभागियों द्वारा दूसरा स्थान प्राप्त किया गया जिसके लिए 30 नवम्बर 2017 को आयोजित नराकास (उ.दि.) की छःमाही बैठक में संस्थान के निदेशक को प्रमाण-पत्र प्रदान किए गए।

संस्थान की वैज्ञानिक, डॉ. अनिदिता दत्ता को भारतीय कृषि अनुसंधान समिति एवं कृषि अनुसंधान संचार केन्द्र द्वारा उनके शोध-पत्र "जर्नलाइज्ड रौ-कॉलम अभिकल्पनाएँ : एक सिंहावलोकन" के लिए उन्हें जून 2017 में "कृषि विज्ञान गौरव" सम्मान से पुरस्कृत किया गया।

संस्थान की वैज्ञानिक, डॉ. अनु शर्मा को राष्ट्रीय प्रौद्योगिकी संस्थान, कुरुक्षेत्र में 22 अगस्त 2017 को

आयोजित भारतीय भाषाओं में प्रथम विज्ञान संगोष्ठी में उनके द्वारा प्रस्तुत शोध-पत्र "सिमेंटिक्स और सॉफ्टवेयर एजेंट आधारित वेब पर्सनलाइज्ड सूचना बहाली" को सर्वश्रेष्ठ शोध-पत्र प्रस्तुति से सम्मानित किया गया।

संस्थान में प्रशासनिक कार्य के साथ-साथ वैज्ञानिक प्रकृति के कार्यों में भी हिन्दी का उपयोग हो रहा है। संस्थान के वैज्ञानिक प्रभागों द्वारा आयोजित प्रशिक्षण कार्यक्रमों की संदर्भ पुस्तिकाओं में कवर पेज, आमुख एवं प्राक्कथन द्विभाषी रूप में प्रस्तुत करने के साथ-साथ कुछ हिन्दी के व्याख्यान भी शामिल किये गये। वैज्ञानिकों द्वारा अपनी परियोजना रिपोर्टों में कवर पेज, आमुख, प्राक्कथन एवं सारांश द्विभाषी रूप में प्रस्तुत किये गये तथा कुछ वैज्ञानिकों द्वारा अपनी परियोजना रिपोर्टों में विषय-सूची एवं तालिकाएँ भी द्विभाषी रूप में प्रस्तुत की गयीं। संस्थान के वैज्ञानिकों द्वारा हिन्दी में वैज्ञानिक विषयों पर हिन्दी कार्यशालाओं का आयोजन किया गया। इसके अतिरिक्त, संस्थान में एम.एससी. तथा पीएच.डी. के विद्यार्थियों द्वारा अपने शोध-प्रबन्धों में सार द्विभाषी रूप में प्रस्तुत किये गये। वैज्ञानिकों एवं तकनीकी कर्मियों द्वारा शोध-पत्र हिन्दी में प्रकाशित किये गये।

गृह मंत्रालय, राजभाषा विभाग से प्राप्त 29 फरवरी 2016 के कार्यालय ज्ञापन सं. 12019/81/2015- रा. भा. (का-2)/पार्ट-2 में निहित दिशानिर्देश कि दो वर्ष के निर्धारित लक्ष्य में संस्थान के समस्त अधिकारियों/कर्मचारियों को कम से कम एक बार हिन्दी कार्यशाला में सहभागिता करने का अवसर मिलना अनिवार्य है। उक्त दिशानिर्देशों के अनुपालन के उद्देश्य से संस्थान में सितम्बर 2016 से मार्च 2018 तक अनेक हिन्दी कार्यशालाएँ आयोजित की गयीं जिनमें अभी तक 60 वैज्ञानिकों, 45 तकनीकी कर्मियों तथा 43



प्रशासनिक कर्मियों द्वारा सहभागिता की जा चुकी है। शेष बचे अधिकारियों/ कर्मचारियों द्वारा शीघ्र ही हिन्दी कार्यशालाओं में सहभागिता कर निर्धारित लक्ष्य को पूरा करने की आशा है।

इसी क्रम में, प्रतिवेदनाधीन अवधि के दौरान संस्थान के विभिन्न वर्ग के कर्मियों के लिए सात हिन्दी कार्यशालाएँ आयोजित की गयीं जिनमें से छः कार्यशालाएँ पूर्णतः वैज्ञानिक विषयों पर थीं। पहली कार्यशाला 24 से 26 मई 2017 के दौरान कृषि जैव सूचना केन्द्र के वैज्ञानिक, डॉ. द्विजेश चन्द्र मिश्र एवं श्री नीरज बुढलाकोटी द्वारा "जैव सूचना विज्ञान : एक परिचय" विषय पर आयोजित की गयी जिसमें 08 वक्ताओं द्वारा विषय से सम्बन्धित 11 उप-विषयों पर व्याख्यान दिये गये। इस कार्यशाला में 13 अधिकारियों द्वारा सहभागिता की गयी। दूसरी कार्यशाला भी कृषि जैव सूचना केन्द्र के वैज्ञानिक, डॉ. सारिका एवं डॉ. मीर आसिफ इकबाल द्वारा 27 से 29 जून 2017 के दौरान "कृषि में जैव सूचना का उपयोग" विषय पर आयोजित की गयी। इस कार्यशाला में 09 वक्ताओं द्वारा विषय से सम्बन्धित 11 उप-विषयों पर व्याख्यान दिये गये तथा इस कार्यशाला में 09 अधिकारियों एवं 01 कर्मचारी ने सहभागिता की। तीसरी कार्यशाला प्रतिदर्श सर्वेक्षण प्रभाग के वैज्ञानिक, सुश्री वन्दिता कुमारी चौधरी, श्री दीपक सिंह एवं श्री प्रदीप बसाक द्वारा 25 से 27 सितम्बर 2017 के दौरान "कृषि में प्रतिदर्श सर्वेक्षण तकनीकों का उपयोग" विषय पर आयोजित की गयी। इस कार्यशाला में 11 वक्ताओं द्वारा 11 विभिन्न विषयों पर व्याख्यान दिये गये। इस कार्यशाला में 12 अधिकारियों ने सहभागिता की। चौथी कार्यशाला हिन्दी एकक की प्रभारी एवं वरिष्ठ तकनीकी अधिकारी, सुश्री ऊषा जैन द्वारा 30 अक्टूबर 2017 को 'राजभाषा नियम एवं अनुपालन' विषय पर आयोजित की गयी। इस कार्यशाला में 08 अधिकारियों एवं 30 कर्मचारियों द्वारा सहभागिता की गयी। पाँचवीं कार्यशाला परीक्षण अभिकल्पना प्रभाग के वैज्ञानिक, डॉ. बी.एन मंडल एवं डॉ. सुकान्त दाश द्वारा 30 नवम्बर से 04 दिसम्बर 2017 के दौरान "एस.ए.एस. द्वारा बहुचर सांख्यिकीय विश्लेषण" विषय पर आयोजित की गयी जिसमें 08

वक्ताओं द्वारा विषय से सम्बन्धित 15 उप-विषयों पर व्याख्यान दिये गये। इस कार्यशाला में 14 अधिकारियों एवं 01 कर्मचारी द्वारा सहभागिता की गयी। छठी कार्यशाला संगणक अनुप्रयोग प्रभाग के वैज्ञानिक, श्री पाल सिंह एवं डॉ. सुदीप कुमार द्वारा 09 जनवरी 2018 को "कृषि में संगणक अनुप्रयोग" विषय पर आयोजित की गयी जिसमें 07 वक्ताओं द्वारा 07 विभिन्न विषयों पर व्याख्यान दिये गये। इस कार्यशाला में 18 अधिकारियों एवं 02 कर्मचारी द्वारा सहभागिता की गयी। उक्त वैज्ञानिक विषयों पर आयोजित हिन्दी कार्यशालाओं में अनेक वक्ता हिन्दीतर थे जिन्होंने बड़ी निपुणता के साथ हिन्दी में व्याख्यान दिये। कार्यशालाओं के आयोजकों/वक्ताओं द्वारा प्रतिभागियों को कुछ व्याख्यानों की सामग्री, मैनुअल के रूप में, हिन्दी भाषा में उपलब्ध करायी गयी। सातवीं कार्यशाला विशेष रूप से संस्थान के वैज्ञानिक वर्ग के लिए 20 जनवरी 2018 को "कृषक कल्याण के लिए सांख्यिकी एवं सूचना-विज्ञान का योगदान" विषय पर परीक्षण अभिकल्पना प्रभाग की अध्यक्षा, डॉ. सीमा जग्गी एवं प्रमुख वैज्ञानिक, डॉ. सिनी वरगीस के संयोजन में हिन्दी एकक द्वारा आयोजित की गयी। इस कार्यशाला में आई.सी.एम.आर. के पूर्व अपर महानिदेशक एवं राष्ट्रीय सांख्यिकीय आयोग के सदस्य, डॉ. पदम सिंह, राष्ट्रीय कृषि आर्थिकी नीति अनुसंधान संस्थान (एन.आई.ए.पी.) के निदेशक, डॉ. सुरेश पाल, संस्थान की पूर्व प्रमुख वैज्ञानिक, डॉ. रंजना अग्रवाल तथा संस्थान के पूर्वानुमान एवं कृषि प्रणाली मॉडलिंग प्रभाग के अध्यक्ष, डॉ. कमलेश नारायण सिंह एवं परीक्षण अभिकल्पना प्रभाग के प्रमुख वैज्ञानिक, डॉ. राजेन्द्र प्रसाद द्वारा संस्थान के कार्यक्षेत्र से सम्बन्धित विभिन्न विषयों पर हिन्दी में व्याख्यान दिये गये। इस कार्यशाला में 58 वैज्ञानिकों द्वारा सहभागिता की गयी।

प्रतिवेदनाधीन अवधि में संस्थान में राजभाषा कार्यान्वयन समिति की बैठकें आयोजित की गयीं। इन बैठकों में राजभाषा अधिनियम, 1963 की धारा 3(3) के अनुपालन को सुनिश्चित करने, राजभाषा विभाग द्वारा जारी वार्षिक कार्यक्रम की विभिन्न मदों, राजभाषा विभाग एवं परिषद मुख्यालय से समय-समय पर प्राप्त



निदेशों का अनुपालन सुनिश्चित करने, कार्यशालाओं के नियमित आयोजन, हिन्दी पत्रिका के प्रकाशन, हिन्दी पखवाड़े के आयोजन इत्यादि पर विस्तार से चर्चा हुई।

राजभाषा विभाग द्वारा जारी वार्षिक कार्यक्रम में निहित लक्ष्यों को पूरा करते हुए संस्थान के अधिकारियों/कर्मचारियों द्वारा समस्त पत्राचार हिन्दी में अथवा द्विभाषी रूप में किया गया। संस्थान के विभिन्न वैज्ञानिक प्रभागों तथा प्रशासनिक अनुभागों द्वारा आयोजित बैठकों की कार्यसूची तथा कार्यवृत्त शत-प्रतिशत हिन्दी में अथवा द्विभाषी रूप में जारी किये गये। संस्थान में अपना कार्य शत-प्रतिशत हिन्दी में करने के लिए 12 अनुभागों को विनिर्दिष्ट किया गया है। गृह मंत्रालय, राजभाषा विभाग द्वारा जारी विभिन्न नकद पुरस्कार योजनाएँ संस्थान में लागू हैं तथा संस्थान के कर्मियों ने इन योजनाओं में भाग लिया।

संस्थान में कार्यरत सभी हिन्दीतर अधिकारियों/कर्मचारियों द्वारा हिन्दी ज्ञान सम्बन्धी प्रशिक्षण पूरा किया जा चुका है। आज तक की स्थिति के अनुसार, संस्थान में अब कोई ऐसा हिन्दीतर अधिकारी/कर्मचारी शेष नहीं रह गया है जिसे हिन्दी ज्ञान सम्बन्धी प्रशिक्षण दिया जाना शेष हो।

इसके अतिरिक्त, 'हिन्दी शिक्षण योजना' के अन्तर्गत हिन्दी आशुलिपि एवं हिन्दी टंकण के प्रशिक्षण का लक्ष्य भी संस्थान द्वारा पूरा कर लिया गया है। राजभाषा विभाग से प्राप्त दिशा-निर्देशों के अनुसरण में वर्ग 'घ' से वर्ग 'ग' में गये कर्मियों में से वर्ग 'ग' श्रेणी के लिए निर्धारित शैक्षिक योग्यता रखने वाले कर्मियों को रोस्टरबद्ध कर उन्हें भी केन्द्रीय हिन्दी प्रशिक्षण संस्थान से हिन्दी टंकण का प्रशिक्षण दिलवाया जा चुका है। परीक्षा में पूरक परिणाम प्राप्त/अनुत्तीर्ण 03 कर्मियों द्वारा जनवरी 2018 में पूरक/पुनः टंकण परीक्षा दी गयी है।

संस्थान की वेबसाइट पर उपलब्ध 'हिन्दी सेवा लिंक' में सांख्यिकीय एवं प्रशासनिक शब्दावली के वर्ण क्रमानुसार कुछ शब्द, कुछ द्विभाषी प्रपत्र, दैनिक काम काज के प्रयोग में आने वाली कुछ टिप्पणियाँ, द्विभाषी पदनाम, वाक्यांश इत्यादि सामग्री उपलब्ध है।

अपना दैनिक कार्य हिन्दी में सरलता से करने के लिए संस्थान के कर्मियों द्वारा इस सेवा का उपयोग किया जाता है।

संस्थान द्वारा प्रकाशित हिन्दी पत्रिका, 'सांख्यिकी-विमर्श' के बारहवें अंक का प्रकाशन मार्च 2017 में किया गया। इस पत्रिका में संस्थान के कीर्तिस्तम्भ, सम्बन्धित वर्ष में किये गये अनुसंधानों व अन्य कार्यों के संक्षिप्त विवरण, राजभाषा से सम्बन्धित कार्यों आदि की जानकारी के साथ-साथ कृषि सांख्यिकी, संगणक अनुप्रयोग एवं कृषि जैव सूचना से सम्बन्धित विभिन्न लेखों एवं शोध-पत्रों को भी प्रस्तुत किया जाता है। पाठकों के हिन्दी ज्ञानवर्धन के लिए दैनिक स्मरणीय शब्द-शतक हिन्दी व अँग्रेजी में दिया जाता है।

संस्थान में 01 से 14 सितम्बर 2017 के दौरान हिन्दी पखवाड़े का आयोजन किया गया। दिनांक 01 सितम्बर 2017 को हिन्दी पखवाड़े का उद्घाटन संस्थान के निदेशक, डॉ. अंजनी कुमार चौबे जी द्वारा किया गया। हिन्दी पखवाड़े के उद्घाटन के तत्पश्चात काव्य-पाठ का आयोजन किया गया। हिन्दी पखवाड़े के दौरान 'डॉ. दसोगा सिंह स्मृति व्याख्यान' के साथ-साथ वैज्ञानिक प्रभागों में हिन्दी में सर्वाधिक वैज्ञानिक कार्य करने के लिए प्रभागीय चल-शील्ड तथा काव्य-पाठ, वाद-विवाद, प्रश्न-मंच, अन्ताक्षरी, काव्य-गोष्ठी, डिजिटल हिन्दी शोध-पत्र प्रस्तुति, हिन्दीतर कर्मियों के लिए हिन्दी श्रुतलेख एवं शब्दार्थ लेखन प्रतियोगिताएँ आयोजित की गयीं। प्रश्न-मंच एवं अन्ताक्षरी प्रतियोगिता के संचालकों द्वारा इन प्रतियोगिताओं को ऑडियो विजुअल रूप में प्रस्तुत किया गया जिससे ये प्रतियोगिताएँ अत्यन्त ही रोचक रहीं। सभी प्रतियोगिताओं में छात्रों सहित संस्थान के विभिन्न वर्ग के कर्मियों ने बढ़-चढ़कर हिस्सा लिया। संस्थान में प्रत्येक वर्ष हिन्दी दिवस के अवसर पर डॉ. दसोगा सिंह स्मृति व्याख्यान का आयोजन किया जाता है। इस वर्ष इस कड़ी का छब्बीसवाँ व्याख्यान संस्थान के पूर्व प्रोफेसर (कृषि सांख्यिकी), डॉ. रणधीर सिंह जी द्वारा दिया गया और इस कार्यक्रम की अध्यक्षता आई.सी.एम.आर. के पूर्व अपर महानिदेशक



एवं राष्ट्रीय सांख्यिकीय आयोग के सदस्य, डॉ. पदम सिंह जी द्वारा की गयी। दिनांक 14 सितम्बर 2017 को हिन्दी पखवाड़े के समापन समारोह के अवसर पर इस दौरान आयोजित प्रतियोगिताओं के सफल प्रतियोगियों को पुरस्कृत करने के साथ-साथ वर्ष 2016-17 के दौरान "सरकारी कामकाज मूल रूप से हिन्दी में करने के लिए प्रोत्साहन योजना" के अन्तर्गत भी नकद पुरस्कार प्रदान किये गये। इसके

अतिरिक्त, इस अवसर पर जुलाई 2016 से जून, 2017 तक की अवधि के दौरान संस्थान में आयोजित हिन्दी कार्यशालाओं के वक्ताओं को भी सम्मानित करने के साथ-साथ संस्थान द्वारा प्रकाशित हिन्दी पत्रिका : सांख्यिकी विमर्श 2016-17 के सम्पादक मंडल के सदस्यों को भी प्रशस्ति-पत्र प्रदान किये गये।

(आभार : लेखिका टंकण हेतु श्री खुशियाल, सहायक की आभारी हैं)

द्विभाषी पदनाम

1	Accountant	लेखाकर
2	Accountant General	महा लेखाकार
3	Accounts Officer	लेखा अधिकारी
4	Acting Director	कार्यकारी निदेशक
5	Additional Secretary	अपर सचिव
6	Administrator General	महाप्रशासक
7	Adviser	सलाहकार
8	Agent	अभिकर्ता, एजेंट
9	Allotment Officer	आबंटन अधिकारी
10	Ambassador	राजदूत
11	Analyst	विश्लेषक
12	Appellate Authority	अपील प्राधिकारी
13	Assistant	सहायक
14	Audit Officer	लेखा-परीक्षा अधिकारी
15	Auditor General	महालेखा परीक्षक
16	Caretaker	रखवाल
17	Cashier	रोकड़िया
18	Chairman	सभापति, अध्यक्ष
19	Chancellor	कुलाधिपति
20	Chief Commissioner	मुख्य आयुक्त / कमिश्नर
21	Civil Engineer	सिविल इंजीनियर, सिविल अभियन्ता
22	Clerk	क्लर्क, लिपिक
23	Commissioner	आयुक्त, कमिश्नर
24	Comptroller and Auditor General	नियंत्रक और महालेखा परीक्षक
25	Controller	नियंत्रक
26	Controller General	महानियंत्रक
27	Controlling Officer	नियंत्रण अधिकारी

28	Convener	संयोजक
29	Coordinating Officer	समन्वय अधिकारी
30	Despatcher	प्रेषक
31	Diarist	डायरी लेखक
32	Director General	महानिदेशक
33	Economic Adviser	आर्थिक सलाहकार
34	Editor	सम्पादक
35	Educational Adviser	शिक्षा सलाहकार
36	Efficiency Officer	कार्यकुशलता अधिकारी
37	Election Commissioner	निर्वाचन आयुक्त
38	Emergency Officer	आपात अधिकारी
39	Enumerator	गणनाकार
40	Establishment Officer	स्थापना अधिकारी
41	Estate Officer	संपदा अधिकारी
42	Evaluation Officer	मूल्यांकन अधिकारी
43	Excise Commissioner	उत्पाद शुल्क आयुक्त
44	Executive Engineer	कार्यपालक इंजीनियर
45	Financial Adviser	वित्त सलाहकार
46	Gazetted Officer	राजपत्रित अधिकारी
47	General Manager	प्रधान प्रबन्धक, महा प्रबन्धक
48	Governor	राज्यपाल
49	Implementation Officer	परिपालन / कार्यान्वयन अधिकारी
50	Income Tax Officer	आयकर अधिकारी
51	Information Officer	सूचना अधिकारी
52	Inspector	निरीक्षक
53	Inspector General of police	पुलिस महानिरीक्षक
54	Investigator	अन्वेषक
55	Judge	न्यायाधीश



56	Judicial Officer	न्यायिक अधिकारी
57	Legal Adviser	कानूनी / विधि सलाहकार
58	Liasion Officer	सम्पर्क अधिकारी
59	Librarian	पुस्तकाध्यक्ष
60	Lower Division Clerk	निम्न श्रेणी लिपिक / क्लर्क
61	Magistrate	मजिस्ट्रेट
62	Manager	प्रबन्धक, मैनेजर
63	Medical Officer	चिकित्सा अधिकारी
64	Officer Incharge	प्रभारी अधिकारी
65	Officer on Special Duty	विशेष कार्य अधिकारी, विशेषाधिकारी
66	Parliamentary Secretary	संसदीय सचिव
67	Personnel Officer	कार्मिक अधिकारी
68	Planning Officer	योजना अधिकारी
69	Police Inspector	पुलिस निरीक्षक
70	Private Secretary	निजी सचिव
71	Pro-chancellor	सम-कुलाधिपति
72	Production Manager	उत्पादन मैनेजर
73	Professor	आचार्य, प्रोफेसर
74	Public Relation Officer	जन सम्पर्क अधिकारी
75	Publication Officer	प्रकाशन अधिकारी
76	Regional Officer	क्षेत्रीय अधिकारी
77	Registrar (of University)	कुल-सचिव
78	Research Assistant	अनुसंधान सहायक

79	Returning Officer	निर्वाचन अधिकारी
80	Sales Tax Officer	बिक्रीकर अधिकारी
81	Secretary	सचिव
82	Secretary General	महा सचिव
83	Section Officer	अनुभाग अधिकारी
84	Security Officer	सुरक्षा अधिकारी
85	Stenographer	आशुलिपिक, स्टेनोग्राफर
86	Sub-Divisional Officer	उपमंडल-अधिकारी
87	Superintendent	अधीक्षक
88	Supervisor	पर्यवेक्षक, सुपरवाइजर
89	Surveyor	सर्वेक्षक
90	Taxation Inspector	कराधान निरीक्षक
91	Translator	अनुवादक
92	Treasurer	कोषपाल, कोषाध्यक्ष
93	Under Secretary	अवर सचिव
94	Upper Division Clerk	उच्च श्रेणी लिपिक / क्लर्क
95	Verification Officer	सत्यापन अधिकारी
96	Vice Chancellor	कुलपति
97	Vice-President	उपराष्ट्रपति, उपप्रधान, उपाध्यक्ष
98	Vigilance Officer	सतर्कता अधिकारी
99	Warden	वार्डन
100	Welfare Officer	कल्याण अधिकारी

(संकलनकर्ता : रुषा जैन)
स्रोत : कार्यालय सहायिका

लैटिन शब्दों के हिन्दी समानक

1	Ad-hoc	तदर्थ
2	Ad interim	अंतःकालीन/अस्थाई
3	Anti dated	पूर्व दिनांकित
4	Bonafide	वास्तविक/असली/सद्भाव से
5	Cum	-व-, और एवं
6	Data	आधार/तथ्य/आंकड़े
7	De facto	वस्तुतः/यथार्थतः
8	De jure	विधितः
9	Ditto	यथोपरि/जैसे ऊपर
10	En route	मार्ग/रास्ते में
11	Errata	शुद्धि-पत्र
12	Ex cadre	संवर्ग बाह्य
13	Ex gratia	अनुग्रहपूर्वक
14	Ex officio	पदेन
15	Ex parte	एक पक्षीय/एकतरफा
16	Ex post facto	कार्योत्तर
17	In absentia	अनुपस्थिति में
18	In lieu of	के बदले में/के स्थान पर
19	In toto	संपूर्णतः/पूरी तरह से
20	Inter alia	अन्य बातों के साथ/ साथ-साथ
21	Inter se	परस्पर/आपस में
22	Liaison	संपर्क
23	Malafide	कदाशय/बदनीयत
24	Modus operandi	कार्य प्रणाली
25	Mutatis mutandis	यथावश्यक परिवर्तन सहित

26	Nota Bene (N.B.)	विशेष ध्यान दीजिए
27	Onus	भार/दायित्व
28	Par excellence	श्रेष्ठ/उत्कृष्ट
29	Per annum	प्रतिवर्ष/वार्षिक
30	Per capita	व्यक्तिवार/प्रति व्यक्ति
31	Percent	प्रतिशत/प्रति सैकड़ा
32	Per se	स्वतः
33	Post script (P.S.)	पुनश्च
34	Prima facie	प्रथम दृष्टया/प्रत्यक्षतः/ प्रथम दृष्टि में
35	Proforma	प्रारूप/प्रपत्र/प्रोफार्म
36	Pro rata	अनुपाततः/यथानुपात
37	Proxy	प्रतिपत्र/प्रतिपत्री/परोक्षी
38	Quantum	प्रमात्रा/क्वांटम
39	Quorum	कोरम
40	Resume	सारांश/सारवृत्त
41	Sine die	अनिश्चित काल के लिए
42	Status quo	यथापूर्व स्थिति
43	Sub-judice	न्यायाधीन
44	Tenure	अवधि/कार्यकाल
45	Veto	निषेधाधिकार/वीटो
46	Vice versa	प्रतिलोमतः
47	Vis-a-vis	आमने-सामने/सम्मुखीन
48	Writ	रिट/आदेश
49	Viva voce	मौखिक परीक्षा
50	Verbatim	शब्दशः

(संकलनकर्ता : ऊषा जैन)
स्रोत : राजभाषा सहायिका

दैनिक स्मरणीय शब्द-शतक

1	Absolute density	निरपेक्ष घनत्व
2	Absolute efficiency	निरपेक्ष दक्षता
3	Absolute humidity	निरपेक्ष आर्द्रता
4	Acceleration coefficient	त्वरण गुणांक
5	Acceptance test	स्वीकारी परीक्षण
6	Accumulated	संचित
7	Accuracy	यथार्थता
8	Agricultural region	कृषि-प्रदेश
9	Agricultural sampling	कृषीय प्रतिचयन
10	Asymmetrical factorial designs	असममित बहु-उपादानीय अभिकल्पनाएँ
11	Asymmetry value	असममित मान
12	Asymptotic	उपगामी
13	Bootstrap technique	बूट-स्ट्रैप तकनीक
14	Calibration	अंशाकन
15	Centre of inversion	प्रतिलोम केन्द्र
16	Characteristic matrix	अभिलक्षणिक आव्यूह
17	Column diagram	स्तम्भ आरेख
18	Column matrix	स्तम्भ आव्यूह
19	Compactness	संहतता
20	Comparability	तुलनात्मक, तुलनीयता
21	Comparative table	तुलनात्मक सारणी
22	Complement	पूरक
23	Complementary function	पूरक फलन
24	Complex data	जटिल आँकड़े
25	Composite function	संयुक्त फलन
26	Composite hypothesis	संयुक्त परिकल्पना
27	Conditioned response	प्रतिबंधित अनुक्रिया

28	Continental climate	महाद्वीपीय जलवायु
29	Correlated samples	सहसम्बन्धित प्रतिदर्श
30	Covariant component	सहपरिवर्ती घटक
31	Deficiency index	न्यूनता सूचकांक
32	Demand equation	माँग समीकरण
33	Determining variable	निर्धारक चर
34	Diagonal point	विकर्ण बिन्दु
35	Diagonal regression	विकर्ण समाश्रयण
36	Exogenous variable	बाह्य चर
37	Expected frequency	प्रत्याशित बारंबारता
38	Expected model	व्याख्यात्मक निदर्श
39	Explicit model	स्पष्ट निदर्श
40	Exponential	घरघातांकी
41	External variable	बाह्य चर
42	Face value	अंकित मूल्य
43	Factorial method	क्रमगुणित विधि
44	Feasible solution	सुसंगत हल
45	Germicide	रोगाणुनाशी
46	Germplasm	जर्मप्लाज्म, जनन द्रव्य
47	Heterogeneity test	विषमांगता परीक्षण
48	Heterogeneous	विषमांगी, विषमजातीय
49	Homogeneity test	समांगता परीक्षण
50	Hypothetical element	परिकल्पित तत्व
51	Inadequacy	अपर्याप्तता
52	Inconsistency	असंगतता
53	Independent function	स्वतंत्र फलन
54	Indirect sampling	परोक्ष प्रतिचयन
55	Lag effect	पश्चता प्रभाव
56	Lagged variable	पश्चायित चर
57	Lag regression	पश्चता-समाश्रयण
58	Land utilization	भूमि-उपयोग



59	Large sample test	बृहत् प्रतिदर्श परीक्षण
60	Linear autoregression	रेखिक स्वसमाश्रयण
61	Linear expansion	रेखीय प्रसार
62	Linear model	रेखिक निदर्श
63	Linear parameter	रेखिक प्राचल
64	Measure of efficiency	दक्षता माप
65	Measure of kurtosis	ककुदता-माप
66	Measure of precision	परिशुद्धता माप
67	Measure of skewness	वैषम्य माप
68	Median test	माधिका-परीक्षण
69	Method of least squares	न्यूनतम वर्ग विधि
70	Multiphase sampling	बहुचरण प्रतिचयन
71	multiple integration	बहु समाकलन
72	multiple regression analysis	बहुसमाश्रयण विश्लेषण
73	Nondimensional equation	अविम समीकरण
74	Nondimensional parameter	अविम प्राचल
75	Normality	प्रसामान्यीकृत
76	Normalized value	प्रसामान्यीकृत मान
77	Observation equation	प्रेषण समीकरण
78	observed error	प्रेषण त्रुटि
79	Optimal decision	इष्टतम निर्णय

80	Optimal estimate	इष्टतम आकल
81	Optimum stratification	इष्टतम स्तरण
82	Optimum weight	इष्टतम भार
83	Organizational structure	संगठनात्मक संरचना
84	Organizational function	लांबिक फलन
85	Orthogonal matrix	लांबिक आव्यूह
86	Partial derivative	आंशिक अवकलन
87	Partial integration	आंशिक समाकलन
88	Performance test	निष्पादन परीक्षण
89	Population	जनसंख्या, संख्या, आबादी, समष्टि
90	Population mean	समष्टि माध्य
91	Positive correlation	धनात्मक सहसम्बन्ध
92	Preference function	वरीयता फलन
93	Probability factor	प्रायिकता गुणांक
94	Projected	प्रक्षेपित
95	Random variable	यादृच्छिक चर
96	Sequential sampling	अनुक्रमिक प्रतिचयन
97	Stability coefficient	स्थायित्व गुणांक
98	Statistical hypothesis	सांख्यिकीय परिकल्पना
99	Two dimensional	द्विविम
100	Under estimation	अवआकलन, अल्पानुमान

(संकलनकर्ता : ऊषा जैन)

स्रोत : बृहत् पारिभाषिक शब्द-संग्रह

विज्ञान : खण्ड 1 एवं 2

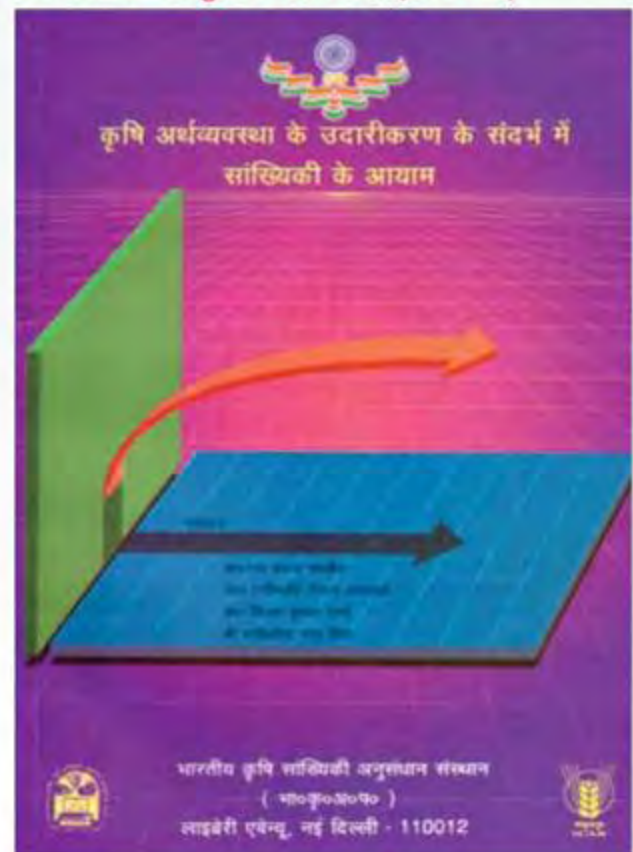
अतीत के झरोखे से...



संस्थान द्वारा प्रथम हिन्दी पत्रिका का
प्रकाशन : 1980-81



भारतीय कृषि सांख्यिकी अनुसंधान संस्थान द्वारा 10 अगस्त 1998 को राजस्थान कृषि महाविद्यालय, उदयपुर में कृषि अर्थव्यवस्था के उदारीकरण के संदर्भ में सांख्यिकी के आयाम विषय पर आयोजित राष्ट्रीय संगोष्ठी में प्रस्तुत शोध-पत्रों पर आधारित संस्तुतियों का संग्रह (प्रथम बार)

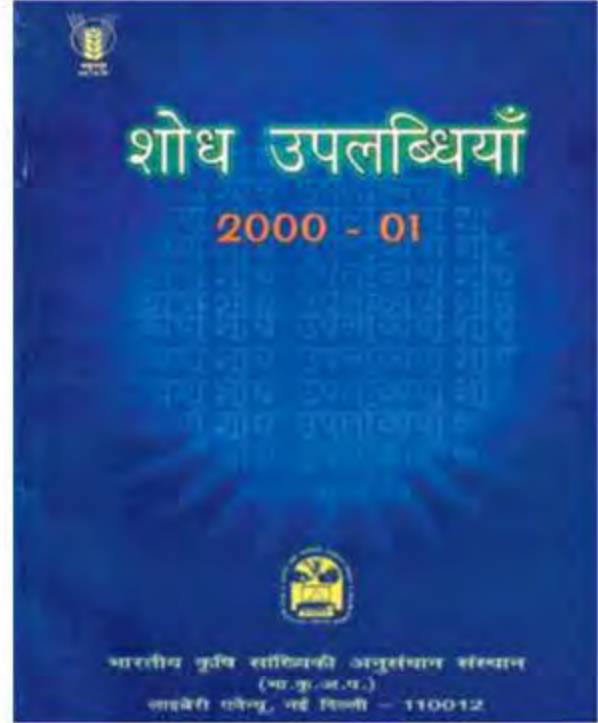




भारतीय कृषि अनुसंधान संस्थान, नई दिल्ली के पी.जी. स्कूल से वर्ष 2001-02 के दौरान संस्थान के वैज्ञानिकों के साथ पीएच.डी. एवं एम.एससी. कर रहे छात्रों द्वारा प्रस्तुत शोध-प्रबन्धों के सार (एबस्ट्रैक्ट्स) के हिन्दी रूपान्तर का शोध बुलेटिन के रूप में संग्रह (प्रथम बार)



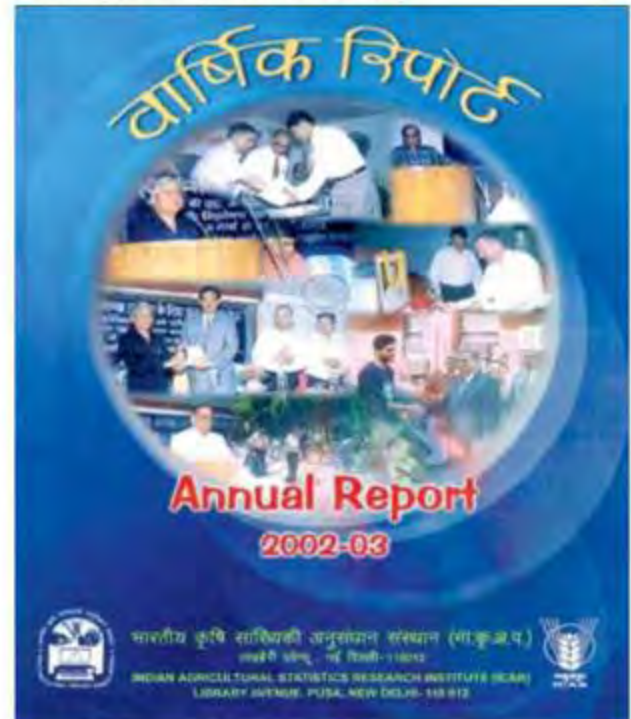
वर्ष 2000-2001 के दौरान संस्थान में चलाई गयी अनुसंधान परियोजनाओं की वर्षभर की उपलब्धियों का हिन्दी में विवरण (प्रथम बार)



संस्थान द्वारा वर्ष 2005-06 में हिन्दी पत्रिका "सांख्यिकी-विमर्श" के प्रथम अंक का प्रकाशन किया गया। परिषद् की "गणेश शंकर विद्यार्थी हिन्दी पत्रिका पुरस्कार" योजना के अन्तर्गत वर्ष 2005-06 में प्रकाशित संस्थान की हिन्दी पत्रिका "सांख्यिकी-विमर्श" को वर्ष 2006 में द्वितीय पुरस्कार से सम्मानित किया गया।



वर्ष 2002-2003 में संस्थान द्वारा अपनी वार्षिक रिपोर्ट द्विभाषी (डिग्लॉट रूप में) प्रकाशित की गयी। वार्षिक रिपोर्ट डिग्लॉट रूप में प्रकाशित करने में भा.कृ.अनु. प्रणाली में संस्थान अग्रणी रहा है तथा इसके लिए संस्थान को परिषद् से प्रशंसा-पत्र प्राप्त हुआ है। (प्रथम बार)





भारतीय कृषि अनुसंधान परिषद की "राजर्षि टण्डन राजभाषा पुरस्कार योजना" के अन्तर्गत वर्ष 2005-06 के दौरान सरकारी कामकाज में हिन्दी के प्रयोग में उल्लेखनीय योगदान के लिए संस्थान को प्रथम पुरस्कार से सम्मानित किया गया ।



भारतीय कृषि अनुसंधान परिषद की "गणेश शंकर विद्यार्थी हिन्दी पत्रिका पुरस्कार योजना" के अन्तर्गत संस्थान की हिन्दी पत्रिका सांख्यिकी विमर्श : 2005-06 को द्वितीय पुरस्कार से सम्मानित किया गया ।



भारतीय कृषि अनुसंधान परिषद की "राजर्षि टण्डन राजभाषा पुरस्कार योजना" के अन्तर्गत वर्ष 2003-04 के दौरान सरकारी कामकाज में हिन्दी के प्रयोग में उल्लेखनीय योगदान के लिए संस्थान को द्वितीय पुरस्कार से सम्मानित किया गया ।



“भारतीय कृषि अनुसंधान परिषद् की “राजर्षि टण्डन राजभाषा पुरस्कार योजना” के अन्तर्गत वर्ष 2003-2004 के दौरान सरकारी कामकाज में उल्लेखनीय योगदान के लिए बड़े संस्थानों के वर्ग में संस्थान को प्राप्त द्वितीय पुरस्कार के सम्मान में जारी प्रमाण-पत्र



“भारतीय कृषि अनुसंधान परिषद् की “राजर्षि टण्डन राजभाषा पुरस्कार योजना” के अन्तर्गत वर्ष 2005-2006 के दौरान सरकारी कामकाज में उल्लेखनीय योगदान के लिए बड़े संस्थानों के वर्ग में संस्थान को प्राप्त प्रथम पुरस्कार के लिए संस्थान को जारी प्रशस्ति-पत्र के सम्मान में जारी प्रमाण-पत्र



“भारतीय कृषि अनुसंधान परिषद की “राजर्षि टण्डन राजभाषा पुरस्कार योजना” के अन्तर्गत वर्ष 2005-2006 के दौरान हिन्दी पत्रिका “सांख्यिकी-विमर्श” प्राप्त द्वितीय पुरस्कार के सम्मान में जारी प्रमाण-पत्र

भारतीय कृषि अनुसंधान परिषद

गणेश शंकर विद्यार्थी हिन्दी पत्रिका पुरस्कार

प्रशस्ति-पत्र

वर्ष 2005-06 के दौरान
माननीय कृषि सांख्यिकी अनुसंधान संस्थान, पूना, नई दिल्ली
द्वारा प्रकाशित हिन्दी पत्रिका
“सांख्यिकी-विमर्श”
को द्वितीय पुरस्कार से सम्मानित किया जाता है।

नई दिल्ली
28 नवम्बर, 2006




राजभाषा सम्मान

भारत सरकार, राजभाषा विभाग की नगर राजभाषा कार्यान्वयन समिति (उत्तरी दिल्ली) द्वारा वर्ष 2015-16 में राजभाषा कार्यान्वयन कार्य में उत्कृष्ट निष्पादन हेतु बड़े कार्यालय वर्ग में भारतीय कृषि सांख्यिकी अनुसंधान संस्थान को तृतीय पुरस्कार प्रदान किया गया। इसके लिए संस्थान को एक ट्रॉफी एवं प्रशस्ति-पत्र प्रदान किया गया।



वर्ष 2015-16 में राजभाषा कार्यान्वयन कार्य में उत्कृष्ट निष्पादन हेतु बड़े कार्यालय वर्ग में भारत सरकार, राजभाषा विभाग की नगर (उत्तरी दिल्ली) द्वारा संस्थान को तृतीय पुरस्कार

हिन्दी पखवाड़ा-2017

उद्घाटन एवं काव्य-पाठ



अन्ताक्षरी



वाद-विवाद



हिन्दी शोध-पत्र प्रस्तुति प्रतियोगिता



प्रश्न-मंच



प्रभागीय चल-शील्ड



काव्य-गोष्ठी



हिन्दी दिवस एवं डॉ दरोगा सिंह स्मृति व्याख्यान का आयोजन



डॉ दरोगा सिंह जी को पुष्पांजलि देते हुए निदेशक



डॉ दरोगा सिंह जी को पुष्पांजलि देते हुए उनके पुत्र



डॉ दरोगा सिंह जी को पुष्पांजलि देते हुए व्याख्यानदाता



निदेशक सम्बोधन करते हुए



व्याख्यानदाता, व्याख्यान देते हुए



डॉ. दरोगा सिंह स्मृति व्याख्यान के अवसर पर संस्थान के निदेशक मुख्य अतिथि का सम्मान करते हुए



मुख्य अतिथि सम्बोधन करते हुए



डॉ. दरोगा सिंह स्मृति व्याख्यान के अवसर पर मुख्य अतिथि व्याख्यानदाता का सम्मान करते हुए



डॉ. दरोगा सिंह स्मृति व्याख्यान के अवसर पर संस्थान के निदेशक डॉ. दरोगा सिंह जी के पुत्र का सम्मान करते हुए



डॉ. दरोगा सिंह जी के पुत्र सम्बोधन करते हुए



पुरस्कार वितरण



शिक्षक दिवस समारोह





नगर राजभाषा कार्यान्वयन समिति (उत्तरी दिल्ली) के सदस्य
कार्यालयों के लिए 30 अगस्त, 2017 को संस्थान में आयोजित प्रश्नोत्तरी प्रतियोगिता



वर्ष 2017-2018 के दौरान संस्थान में हिन्दी कार्यशालाओं का आयोजन

24-26 मई 2017 के दौरान



27 से 29 जून 2017 के दौरान





25 से 27 सितम्बर 2017 के दौरान



30 अक्टूबर 2017





30 नवम्बर से 04 दिसम्बर 2017





09 जनवरी 2018 को आयोजित



20 जनवरी 2018 को आयोजित हिन्दी कार्यशाला





ऑर्गेनोग्राम

अनुसंधान सलाहकार समिति

निदेशक

संस्थान प्रबंधन समिति

प्रभागाध्यक्ष

अधिकारी प्रभारी

प्रोफे. (कृषि सांख्यिकी)
प्रोफे. (संगणक अनुप्रयोग)
प्रोफे. (जैवसूचना विज्ञान)
बोर्डन

सतर्कता अधिकारी

वरिष्ठ प्रशासनिक अधिकारी

वरि. वित्त एवं लेखाधिकारी

- परीक्षण अभिकल्पना
- सांख्यिकी आनुवंशिकी
- पूर्वानुमान एवं कृषि प्रणाली मॉडलिंग
- प्रतिदर्श सर्वेक्षण
- संगणक अनुप्रयोग
- कृषि जैव-सूचना केंद्र

- पुस्तकालय एवं प्रलेखन
- उन्नत संकाय प्रशिक्षण केंद्र
- संस्थान प्रौद्योगिकी प्रबंधन एकक
- प्राथमिकीकरण, निगरानी एवं मूल्यांकन प्रकोष्ठ
- कृषिज्ञान प्रबंधन प्रकोष्ठ
- एनएसएसएम
- गेस्ट हाउस/आईटीएच

ओएमबी एकक

प्रशिक्षण प्रशासन प्रकोष्ठ

- प्रशा. I अनुभाग
- प्रशा. II अनुभाग
- रोकड़ अनुभाग
- केंद्रीय खरीद अनुभाग
- पावती एवं प्रेषण अनुभाग
- हिन्दी एकक
- कार्य अनुभाग
- रखरखाव अनुभाग
- भण्डार एकक
- उपकरण रखरखाव एकक

- लेखा-परीक्षा अनुभाग
- लेखा अनुभाग



केविके पोर्टल एवं केविके मोबाइल ऐप



कृषि विज्ञान केंद्र ज्ञान सत्र ऑनलाइन पोर्टल (<https://kvk.icas.gov.in/>) केविके में उपलब्ध सुविधाओं एवं कार्यक्रमों के बारे में किसानों को जानकारी देता है, फसल व अन्य उद्यमों सम्बंधित कार्य-प्रणाली, कृषि उत्पादों के बाजार मूल्य, मौसम से संबंधित कृषि विज्ञान सलाह एवं कृषि क्षेत्र में आकरिमिक परियोजना का प्रावधान करता है। इस पोर्टल के द्वारा सभी केविके, कृषि से सम्बंधित जानकारी समय पर अपलोड कर सकते हैं, जिससे कृषकों को लाभ पहुँच पाता है। किसानों के लिए केविके मोबाइल ऐप भी विकसित किया गया है और यह गूगल प्ले स्टोर में उपलब्ध है (<https://play.google.com/store/apps/details?id=com.icas.iasri.kvkappl&hl=en>)। ऐप से जानकारी प्राप्त करने के लिए किसानों को पंजीकरण और प्राथमिक केविके का चयन करना होगा। किसान कृषि से संबंधित सवाल केविके विशेषज्ञों को भेज कर उत्तरका समाधान प्राप्त कर सकते हैं।



राज्य और जिला चयन कर के केविके के बारे में जानकारी प्राप्त करें



केविके में उपलब्ध सुविधाओं की जानकारी प्राप्त करें



मौसम से संबंधित कृषि विज्ञान सलाह को अंग्रेजी अथवा स्थानीय भाषा में डाउनलोड करें



फसल, मछली पालन, बागवानी और पशुपालन के लिए कार्य-प्रणाली पता करें



eNAM पोर्टल से कृषि उत्पादों के बाजार मूल्य पता करें



केविके द्वारा आयोजित किये गये कार्यक्रमों जैसे किसान मेला, कृषि प्रदर्शनी इत्यादि का विवरण प्राप्त करें



केविके वेब पोर्टल देखें



केविके में आयोजित होने वाले कार्यक्रमों की अधिम जानकारी प्राप्त करें



विशेषज्ञों को खेत से संबंधित सवाल भेजें और उनका समाधान प्राप्त करें



प्राथमिक कृषि विज्ञान केंद्र बदलें

कृषि विस्तार सभाग-भारतीय कृषि अनुसंधान परिषद, कृषि अनुसंधान भवन-1, नई दिल्ली-110 012 (भारत)



समवाक्य-अनुसंधान प्रभाग

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

लाइब्रेरी एवम्पू, पूसा, नई दिल्ली-110 012 (भारत)



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

लाइब्रेरी एवम्पू, पूसा, नई दिल्ली-110 012

www.iasri.res.in

आईएसओ 9001:2008 प्रमाणित संस्थान

आईएसओ/आईसी 20000 तथा 27001 प्रमाणित डाटा केंद्र