खंड 3 सार्वजनिक स्वास्थ्य में अनुसंधान और सांख्यिकीय विधियां

ignou
THE PEOPLE'S
UNIVERSITY

इकाई 8 अनुसंधान विधियां और सांख्यिकी उपकरण*

इकाई की रूपरेखा

अधिगम के उद्देश्य

इस इकाई को पढ़ने के बाद, आप समझ पाएंगे:

- अनुसंधान के मूल सिद्धांत;
- अनुसंधान में सांख्यिकीय उपकरणों की आवश्यकता;
- निदर्शन अध्ययन और केस अध्ययन के लिए प्रोटोकॉल का प्रारूप;
- आंकड़ों का सांख्यिकीय विवरण; तथा
- सांख्यिकीय परीक्षण का महत्व।

* योगदानकर्ता – प्रो. के.वी.एस. सरमा (सेवानिवृत्त), सांख्यिकी विभाग, श्री वेंकटेश्वर विश्वविद्यालय, तिरुपति.
अनुवादक – डॉ. निशीथ राय, सहायक प्रोफेसर, मानव विज्ञान, म.गां.अ.हिं.वि. वर्धा, महाराष्ट्र।

8.0 परिचय

एक व्यक्ति का स्वास्थ्य, जीवन की गुणवत्ता का प्राथमिक निर्धारक है। लगभग 70% भारतीय आबादी ग्रामीण क्षेत्रों में रहती है। ग्रामीण और शहरी क्षेत्रों के बीच स्वास्थ्य सेवा कि पहुंच में भारी असमानता है। लोगों पर बीमारी का बोझ किसी भी देश में स्वास्थ्य सेवा का एक संकेतक है। विश्व स्वास्थ्य संगठन (डब्ल्यूएचओ) के अनुसार 1990 में कुल मृत्यु का 53.6% संक्रामक रोगों के कारण थी, जिनमें मातृ, नवजात और पोषण संबंधी बीमारियाँ शामिल थीं। 2016 में कैंसर, हृदय रोग, मधुमेह आदि सहित गैर-संचारी रोगों के कारण कुल मौतों का प्रतिशत 61.8% था (स्रोतः हेल्थ ऑफ नेशन स्टेटः भारत राज्य-स्तरीय बीमारी, 2017 भारतीय चिकित्सा परिषद द्वारा प्रकाशित रिसर्च, पब्लिक हेल्थ फाउंडेशन ऑफ इंडिया और इंस्टीट्यूट फॉर हेल्थ मेट्रिक्स एंड इवैल्यूएशन)।

सस्ती दरों पर गुणवत्ता युक्त स्वास्थ्य सेवा एक महत्वपूर्ण मुद्दा है। स्वास्थ्य देखभाल सेवाओं के वित्तीय बोझ को पूरा करने के लिए केंद्र और राज्य सरकारों के साथ-साथ स्वास्थ्य बीमा कंपनियां कई योजनाएं लेकर आई हैं। हालांकि, थायिल और जीजा (Thayyil and Jeeja) द्वारा 2013 में किये गए एक अध्ययन के अनुसार स्वास्थ्य देखभाल वितरण प्रणाली में एक प्रमुख भूमिका निजी क्षेत्र की है, जिसमें देश के 58% अस्पताल, अस्पतालों में 29% बेड और 81% डॉक्टर हैं।

सार्वजनिक स्वास्थ्य संस्थानों का कामकाज काफी हद तक उनके साथ उपलब्ध सांख्यिकीय आंकड़ों पर निर्भर करता है। उदाहरण के लिए, सरकार बिना शौचालय वाले घरों के प्रतिशत, सुरक्षित पेयजल वाले गांवों की संख्या, टीकाकरण किए जाने वाले बच्चों की संख्या आदि जैसे मुद्दों पर आंकडें प्राप्त करती है।

विभिन्न उद्देश्यों के साथ अभिनव स्वास्थ्य देखभाल के क्षेत्र में बहुत सारे शोध चल रहे हैं। कुछ नीचे सूचीबद्ध हैंः

बीमारियों का कारण बनने वाले कारकों की पहचान करने के नए तरीके खोजना;

स्वस्थ जीवन शैली पर लोगों को शिक्षित करना;

कैंसर, गुर्दे की बीमारियों और दिल की बीमारियों जैसे विशिष्ट रोगों के कारण होने वाली मौतों (मृत्यु) की संख्या का अनुमान लगाना;

नैदानिक जांच (जैसे एक्स-रे, स्कैन, एमआरआई आदि) के लिए नई और बेहतर विधियाँ, दवा, रोगी की देखभाल, अनुवर्ती (फॉलो-अप) आदि।

इन सबके लिये एक वैज्ञानिक दृष्टिकोण की जरूरत होती है और अनुसंधान टीम में आमतौर पर एक सांख्यिकीविद् शामिल होता है।

सांख्यिकी की भूमिकाः सांख्यिकी एक ऐसा विषय है जो डेटा के संग्रह, संगठन और विश्लेषण से संबंधित है। यह प्रतिदर्श डेटा के आधार पर जनसंख्या के बारे में निष्कर्ष निकालने में मदद करता है। अनुसंधान अध्ययन को मोटे तौर पर निम्नानुसार वर्गीकृत किया गया हैः

क) संभावित अध्ययन जिसमें परिणामों को हस्तक्षेपों (ज्ञात एंटीसेडेंट्स) के जवाब में देखा जाता है। वे अवलोकन अध्ययन या तुलनात्मक अध्ययन हो सकते हैं।

ख) केस कंट्रोल स्टडी जो प्रकृति में पूर्वव्यापी अध्ययन हैं। परिणाम ज्ञात हैं, और शोधकर्ता परिणाम के संभावित कारणों की जांच करते हैं।

प्रत्येक शोध अध्ययन के लिए एक अच्छी तरह से लिखित प्रोटोकॉल की आवश्यकता होती है जोः क) अध्ययन का उद्देश्य (यों), ख) लक्ष्य समूह (कोहर्ट) को संबोधित किया जाना है, ग) अध्ययन की अवधि, घ) निदर्शन प्रारूप, आंकड़ों की संकलन और विश्लेषण विधि, च) नैतिक मुद्दे, यदि कोई हो, ज) बजट अनुमान आदि।

निम्नलिखित अनुभाग में हम प्रतिदर्श से संबंधित कुछ अवधारणाओं को समझेंगे।

8.1 जनसंख्या और प्रतिदर्श

एक जनसंख्या अध्ययन के लक्ष्य से संबंधित सभी विषयों (लोगों, जानवरों, पौधों आदि) का संग्रह है और इसे कोहोर्ट या अध्ययन समूह के रूप में भी जाना जाता है। एक प्रतिदर्श (सैंपल) जनसंख्या का एक प्रतिनिधि भाग (सबसेट) है।

उदाहरण के लिए, हम जनसंख्या को '30 साल से कम उम्र की सभी एनीमिया से पीड़ित महिलाओं' के रूप में परिभाषित कर सकते हैं। इस जनसंख्या को हम ठीक से नहीं जानते हैं; एक शोधकर्ता जनसंख्या की विशेषताओं को प्रतिदर्श अध्ययन की मदद से समझता है। जनसंख्या का आकार आमतौर पर बड़ा होता है और यदि जनसंख्या के प्रत्येक सदस्य का अध्ययन किया जाना है, तो इसे जनगणना या स्क्रीनिंग कहा जाता है। यह, हालांकि, महंगा है, समय लगता है और आंकड़े संकलन हेतु प्रशिक्षित जांचकर्ताओं की एक बड़ी टीम की मांग करता है। कुछ मामलों में, स्क्रीनिंग का कोई मतलब नहीं है क्योंकि यह ठीक उसी प्रकार होगा जैसे 'रक्त-शर्करा स्तर जानने के लिए किसी व्यक्ति से कुल रक्त खींचने का प्रयास किया जाये'।

दूसरी ओर, निदर्शन कम खर्चीला है और कुछ प्रशिक्षित व्यक्तियों के साथ डेटा का संकलन किया जा सकता है। प्रतिदर्श से प्राप्त परिणामों को जनसंख्या के लिए सामान्यीकृत किया जाता है। इसे आगमनात्मक दृष्टिकोण के रूप में जाना जाता है और परिणाम अक्सर लक्ष्य समूह के अज्ञात मापदंडों के अनुमान के रूप में माना जाता है।

एक प्रतिदर्श प्रारूप एक योजना है जिसके अनुसार अध्ययन में डेटा का संकलन किया जाएगा। यह इस तरह के प्रतिदर्श, प्रतिदर्श फ्रेम (जनसंख्या सदस्यों की पूरी सूची), प्रतिदर्श आकार, प्रतिदर्श की विधि, प्रश्नावली के प्रारूप, डेटा की प्रविष्टि और डेटा के सत्यापन और विश्वसनीयता उपायों के रूप में निर्दिष्ट करता है।

निदर्शन निष्पक्ष होना चाहिए ताकि जांचकर्ता उत्तरदाताओं या डेटा के संग्रह की प्रक्रिया के चयन को प्रभावित न करे। प्रतिदर्श लेने के तरीके दो प्रकार के होते हैं। यादृच्छिक प्रतिदर्श और गैर-यादृच्छिक प्रतिदर्श। हम नीचे इन तरीकों के बारे में समझेंगेः

अपनी प्रगति जांचें

1) स्वास्थ्य अनुसंधान में सांख्यिकी की भूमिका पर एक नोट लिखें।

.....

.....

.....

.....

.....

2) उपयुक्त उदाहरणों से जनसंख्या और प्रतिदर्श के बीच अंतर लिखिए।

.....

.....

.....

.....

.....

3) एक प्रतिदर्श प्रारूप क्या है? इसके मुख्य घटक क्या हैं?

.....

.....

.....

.....

.....

8.2 यादृच्छिक (रैंडम) निदर्शन

इस विधि में, जनसंख्या के सदस्यों (इकाइयों) को यादृच्छिक या लॉटरी विधि द्वारा प्रतिदर्श में शामिल करते हैं। यह व्यक्तिगत पूर्वाग्रह को सदस्यों को अध्ययन में शामिल करने से रोकता है। यह निष्पक्ष निष्कर्ष निकालने का सबसे अच्छा तरीका है। हम यादृच्छिक निदर्शन के निम्नलिखित तरीके हैं।

8.2.1 सरल यादृच्छिक (रैंडम) निदर्शन

इस विधि में जनसंख्या की प्रत्येक इकाई को प्रतिदर्श में चयनित होने की समान संभावना होती है। यह तब लागू होता है जब जनसंख्या उम्र, लिंग, शरीर द्रव्यमान, शिक्षा के स्तर आदि जैसे कारकों के संबंध में सजातीय होती है। उदाहरण के लिए, एक गाँव में 150 घर हैं। 30 घरों का चयन करने के लिए, 150 पर्ची लिखें, प्रत्येक पर्ची में घर का नंबर हो, उन्हें एक बॉक्स में रखें, पर्चियों को फेंटे और एक समय में एक पर्ची का चयन करें, जब तक कि 30 घरों का चयन नहीं किया जाता है (पुनरावृत्ति को रोक दिया जाता है)।

8.2.2 स्तरीकृत यादृच्छिक निदर्शन

इस पद्धति का उपयोग तब किया जाता है जब जनसंख्या की इकाइयां विषम होती हैं जैसे कि शिक्षा के स्तर, निवास स्थान, सामाजिक आर्थिक स्थिति आदि में विषमता होना। प्रत्येक समूह को एक स्तर कहा जाता है और निदर्शन को कुछ समूह के अधिक प्रतिनिधित्व से बचने के लिए सभी स्तर से सामान प्रतिदर्श लेना चाहिए।

8.2.3 क्रमबद्ध निदर्शन

इस पद्धति का उपयोग तब किया जाता है जब आबादी इकाइयों को पहले से ही एक क्रम में व्यवस्थित किया जाता है जैसे कि कॉलोनी में आवासीय मकान जैसे 1, 2, 3, 3, 100। क्रमबद्ध निदर्शन यादृच्छिक रूप से एक इकाई से शुरू होता है और मकानों के निश्चित अंतराल के साथ क्रमिक मकानों का चयन करता है।

8.2.4 गुच्छ (क्लस्टर) निदर्शन

गुच्छ (क्लस्टर) निदर्शन त्वरित और आसान है। मान लीजिए कि हम टीकाकरण पर एक अध्ययन करना चाहते हैं। उदाहरण के लिए, हम 20 गाँवों का चयन करते हैं, प्रत्येक गाँव एक गुच्छ (क्लस्टर) है । प्रत्येक क्लस्टर के भीतर हम 30 घरों को यादृच्छिक रूप से लेते हैं ताकि 600 घरों को अध्ययन द्वारा कवर किया जाएगा। प्रत्येक क्लस्टर में विषम सदस्य होते हैं और इसलिए जनसंख्या की सबसे अधिक विशेषताओं का प्रतिनिधित्व करता है। यह विश्व स्वास्थ्य संगठन (डब्ल्यूएचओ) द्वारा टीकाकरण पर सर्वेक्षण करने के लिए अनुशंसित विधि है।

8.2.5 बहुस्तरीय (मल्टी स्टेज) निदर्शन

किसी राज्य जैसे बड़े क्षेत्र या राज्य के भीतर के बड़े क्षेत्र में सर्वेक्षण करने के लिए यह तरीका आवश्यक है। मान लें कि हम एक निश्चित संख्या में घरों में जाकर किसी क्षेत्र में एनीमिया की व्यापकता का अध्ययन करना चाहते हैं। फिर पहले चरण में यादृच्छिक आधार पर जिलों की एक पूर्व निर्धारित संख्या का चयन करना सुविधाजनक होता है, यादृच्छिक आधार पर जिलों की एक पूर्व निर्धारित संख्या चुनने के उपरांत हम चरण-2 में कुछ प्राथमिक स्वास्थ्य केंद्रों (PHC) को यादृच्छिक रूप से चुन सकते हैं क्योंकि प्रत्येक PHC कुछ गांवों को कवर करता है। चरण-3 में हम यादृच्छिक आधार पर गांवों की एक निश्चित संख्या का चयन करते हैं और अंत में प्रत्येक गांव में पूर्व निर्धारित संख्या में घरों को यादृच्छिक आधार पर चुना जा सकता है। इस प्रकार, इस विधि में, नमूना इकाइयां एक चरण से दूसरे चरण में बदल जाती हैं। नमूने के आकार के साथ-साथ नमूने की विधि अध्ययन प्रोटोकॉल में निर्दिष्ट की जाति है।

8.3 गैर-यादृच्छिक निदर्शन

गैर-यादृच्छिक निदर्शन (नॉन-रेंडम सैंपलिंग) एक अन्य विधि है, जहां शोधकर्ता उद्देश्यपूर्ण रूप से अध्ययन के लिए प्रासंगिक व्यक्तियों का एक समूह का चयन करता है। यद्यपि यह पद्धति वैज्ञानिक दृष्टिकोण का समर्थन नहीं करती है, फिर भी स्वास्थ्य सेवा या बीमा पर सर्वेक्षण जिसमें त्वरित परिणाम प्राप्त करने हों तो यह उपयोगी होती है। ऐसा सर्वेक्षण आमतौर पर केवल उन लोगों से किया जाता है जिन्होंने बीमा के लिए पंजीकरण किया है और शोधकर्ता के लिए सुलभ हैं। हालांकि, ऐसे अध्ययनों के परिणामों को लक्ष्य समूह (जनसंख्या) के लिए सामान्यीकृत नहीं किया जा सकता है। स्नोबॉल सैंपलिंग एक विशेष चिकित्सा स्थिति वाले लोगों से जानकारी निकालने का एक तरीका है जिसमें वर्तिकाग्र/ किसी रोग का विशेष चिह्न (स्टिग्मा) होता है । उदाहरणों में यौनकर्मियों, ड्रग एब्यूजर्स या एचआईवी रोगियों पर सर्वेक्षण शामिल हैं। ऐसे लोगों की पूरी आबादी को जानना मुश्किल है और इसलिए एक यादृच्छिक नमूना संभव नहीं है। यदि हम अध्ययन के लिए प्रासंगिक एक व्यक्ति को पकड़ने में सक्षम हैं, तो वह अध्ययन क्षेत्र में समान व्यक्तियों तक पहुंचने के लिए एक मार्गदर्शक के रूप में कार्य कर सकता है और ऐसे सभी व्यक्ति प्रतिदर्श बनाते हैं।

8.4 केस नियंत्रित (केस कंट्रोल) अध्ययन

एक केस-कंट्रोल अध्ययन में शोधकर्ता उन रोगियों के एक समूह की जांच करता है जिन्हें स्वास्थ्य की स्थिति के साथ पहचाना जाता है। उदाहरण के लिए, हाइपोथायरायडिज्म वाले व्यक्ति जिसमें शोधकर्ता इसके लिए संभावित कारणों की पहचान करने का प्रयास करता है ।

यह उन मामलों में चयन करने के लिए आम है, जब केवल उन रोगियों के अध्ययन करना हो जो बीमारी के उपचार हेतु नए निदान के लिए जाते हैं। क्योंकि उन स्थितियों में संपर्क की पहचान करना आसान है जो बीमारी का कारण हो सकती हैं। नियंत्रण विषयों को आमतौर पर मिलान के रूप में लिया जाता है, इस अर्थ में कि उनमें अधिकांश कारक (जैसे उम्र, लिंग, शरीर का वजन) सामान होते हैं परंतु वे बीमार नहीं होते हैं ।

मधुमेह, मोटापा, उम्र जैसे कारकों का पैटर्न तब संभव कारणों के रूप में काम करेगा जो अक्सर बीमारी के लिए बायोमार्कर के रूप में जाना जाता है। केस-कंट्रोल अध्ययन में, मामलों की तुलना उन नियंत्रणों से की जाती है जो या तो सामान्य होते हैं या जिनका प्लेसीबो के साथ इलाज किया जाता है। यह कारकों की पहचान करने में मदद करता है, यदि कोई हो, जो मामलों में प्रमुख है, लेकिन नियंत्रण में नहीं है। इंद्रायन और सत्यनारायण (2006) के अनुसंधान डिजाइन के कई व्यावहारिक उदाहरण हैं।

निम्नलिखित अनुभाग में हम सांख्यिकीय डेटा का वर्णन करने के बुनियादी सांख्यिकीय तरीकों का अध्ययन करेंगे।

8.5 वर्णनात्मक सांख्यिकी

सांख्यिकीय डेटा आमतौर पर तीन रूपों अर्थात तालिकाओं, ग्राफ और सारांश मानों में व्यक्त किया जाता है। डेटा जो एक नाममात्र या क्रमिक पैमाने पर बनाया जाता है उनको संख्या (आवृति) और प्रतिशत द्वारा संक्षेपित किया जाता है। ऐसे डेटा को श्रेणीबद्ध डेटा कहा जाता है। यदि डेटा को अंतराल पैमाने पर मापा जाता है, तो हम औसत और भिन्नता के कुछ उपायों का उपयोग करके इसे संक्षेप में प्रस्तुत करते हैं।

आइए हम तालिका 8.1 में श्रेणीबद्ध डेटा और उसके सारांश की एक स्थिति देखें।

तालिका 8.1 (श्रेणीबद्ध डेटा का विवरण): निम्न डेटा कुष्ठ रोगियों और लिंग के प्रकार के अनुसार कुष्ठ रोगियों के वितरण को संदर्भित करता है

कुष्ठ रोग श्रेणी	रोगियों की संख्या		
	पुरुष	महिला	कुल
ट्युबरक्युलॉइड	77	74	151
लेप्रोमेटस	35	33	68
अनिश्चित	10	8	18
सीमावर्ती	7	5	12
कुल	129	120	249

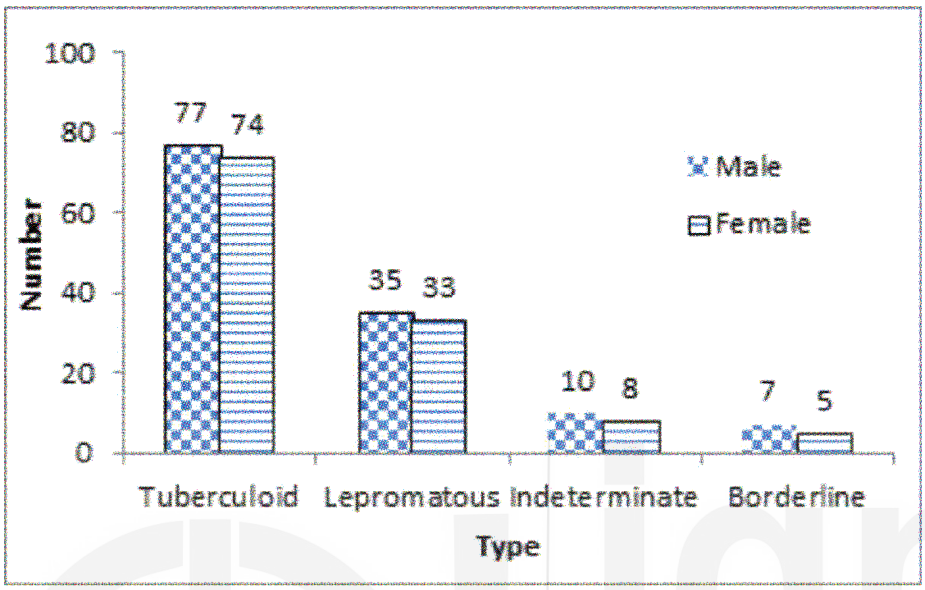
इस डेटा में कुष्ठ रोग के प्रकार और लिंग-वार मामलों की संख्या की जानकारी है।

उपरोक्त तालिका से यह स्पष्ट है कि 249 रोगियों में से 151 (60.64%) को ट्यूबरकुलॉइड है। पुरुषों और महिलाओं के बीच रोग का पैटर्न कमोबेश समान है।

रुचि का चर कुष्ठ रोग का प्रकार है जिसे 4 श्रेणियों के रूप में दिया गया है। प्रत्येक श्रेणी का डेटा गणना (देखे गए मामलों की संख्या) है। लिंग (पुरुष और महिला) एक कारक है। इसलिए डेटा दो आयामी है। इस तरह के डेटा का सारांश आमतौर पर प्रतिशत के संदर्भ में किया जाता है। उदाहरण के लिए, आप देख सकते हैं कि लेप्रोमेटस के कितने प्रतिशत पुरुष सामने आते हैं? इसका उत्तर 249 में से केवल 35 या 14.05% है।

आप देख सकते हैं कि कितने प्रतिशत महिला मरीज सीमावर्ती (बॉर्डरलाइन) श्रेणी में आते हैं। आपको 4.16% मिलना चाहिए। हम यह भी मानते हैं कि ट्यूबरकुलॉइड 249 में से 151 के साथ कुष्ठ रोग का प्रमुख प्रकार है, जिसका अर्थ है 60.6%

इस तरह के डेटा का वर्णन करने का एक अन्य तरीका बार चार्ट का उपयोग है जैसा कि चित्र 8.1 में दिखाया गया है।



चित्र 8.1: कुष्ठ रोग के प्रकार से रोगी का वितरण

एक विकल्प के रूप में आप पुरुष और महिला रोगियों के लिए अलग पाई चार्ट द्वारा डेटा का वर्णन कर सकते हैं। इन चार्ट को बनाने के लिए एक्सेल का उपयोग किया जा सकता है।

मान लीजिए कि आपके पास मिलीग्राम/डीएल में मापा गया सीरम (रक्त) क्रिएटिनिन पर डेटा है। तो, डेटा निरंतर है। मान जरूरी नहीं कि पूरी संख्या हो वह दशमलव में भी हो सकता है ।

इसी तरह, बॉडी मास इंडेक्स, उपवास रक्त ग्लूकोज और एक नवजात बच्चे के जन्म का वजन निरंतर चर के कुछ उदाहरण हैं। ऐसे चरों का औसत वर्णन मायने रखता है बजाय आवृति और प्रतिशत के।

8.6 केंद्रीय प्रवृत्ति के माप

एक डेटा में बहुत बार, हम एक प्रवृत्ति पाते हैं कि अधिकांश डेटा एक केंद्रीय मूल्य के आसपास रहता है जिसे औसत के रूप में जाना जाता है। इसे केंद्रीय प्रवृत्ति कहा जाता है और औसत कहे जाने वाले कुछ मापों के संदर्भ में व्यक्त किया जाता है। हम आमतौर पर इस्तेमाल किए जाने वाले तीन औसत पर चर्चा करेंगे।

8.6.1 माध्य (मीन)

सबसे अधिक इस्तेमाल किया जाने वाला औसत अंकगणित माध्य या केवल माध्य है। यह बस मूल्यों की संख्या से विभाजित सभी मूल्यों का योग है। यदि $x_1, x_2,\ldots, x_n$ डेटा में n-मान हैं, तो माध्य द्वारा दिया जाता है

$$\overline{X} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

उदाहरण 8.2: 13 रोगियों का क्रिएटिनिन स्तर (मिलीग्राम/डीएल) नीचे दिया गया है।

0.70, 0.60, 0.20, 0.26, 0.40, 0.50, 0.35, 0.15, 0.30, 0.36, 0.60, 0.20 और 0.45 इन 13 मानों का योग 5.07 है। तो, आपका माध्य = 5.07 / 13 = 0.39 mg / dl होता है।

माध्य एक महत्वपूर्ण माप है और लोकप्रिय रूप से डेटा के विश्लेषण में उपयोग किया जाता है। यह सभी डेटा मूल्यों पर आधारित है और इसका मान सबसे बड़े मूल्य से बड़ा नहीं हो सकता है। माध्य का अर्थ इस अर्थ में भी है कि डेटा में बहुत बड़े या बहुत छोटे मानों को शामिल करने से मान में अत्यधिक परिवर्तन हो जाता है।

आप ध्यान दे सकते हैं कि मानों (औसत) का मान {2, 3, 4, 5, 6} 20/5 = 4 है। मान लीजिए कि अंतिम मान गलती से 36 के रूप में दर्ज किया गया है; मतलब 50/5 = 10 पर जाता है! फिर से, मूल डेटा में जब 6 को 0 के रूप में लिखा जाता है, तो इसका मतलब 14/5 = 2.8 हो जाता है (मानों की संख्या अभी भी 5 है क्योंकि 0 भी एक मान है)।

जब डेटा का आकार बड़ा होता है तो 100 या 300 मान कहते हैं, डेटा का वर्णन करने का एक तरीका आवृत्ति तालिका या गणना की तालिका तैयार करके है। इसका एक उदाहरण है।

तालिका 8.3: किसी जिले के विभिन्न हिस्सों से एकत्र किए गए 200 नमूनों में देखे गए फ्लोराइड स्तर (मिलीग्राम/एल) को तालिका में दिया गया है।

फ्लोराइड स्तर	0.3 - 0.5	0.5-0.7	0.7-0.9	0.9-1.1	1.1-1.3	1.3-1.5	1.5-1.7
#प्रतिदर्श	6	24	67	61	22	12	8

संख्या या आवृत्ति इंगित करता है।

हम औसत (माध्य) फ्लोराइड स्तर जानना चाहते हैं।

विश्लेषण: हमने फ्लोराइड स्तर के अंतराल का उपयोग करके डेटा का वर्णन किया है और उन नमूनों की संख्या गिनी है जिनके लिए मूल्य प्रत्येक अंतराल का है। इन अंतरालों को वर्ग या बिन भी कहा जाता है। प्रत्येक अंतराल में, ऊपरी सीमा से कम सभी मूल्यों को गिना जाएगा। अब इस डेटा का मतलब खोजने के लिए, हम प्रत्येक अंतराल का मध्य मान निकालते हैं जो ऊपरी और निचली सीमाओं का औसत है। उदाहरण के लिए, अंतराल का मध्य मान 0.7-0.9 (0.7+0.9)/2 = 0.8 है। तब माध्य इस प्रकार पाया जाता है (* इंगित करता है गुणा)।

{0.4*6 + 0.6 *24+0.8 * 67+1.0 * 61+1.2 * 22+1.4 * 12+ 1.6X8}/200

यह = 187.4/200 = 0.94 मिलीग्राम/एल देता है। इसका मतलब है कि अध्ययन क्षेत्र में फ्लोराइड का औसतन स्तर 0.94 mg/L है।

8.6.2 माध्यिका (मीडियन)

माध्यका (मीडियन) एक और औसत है जो अक्सर क्रमिक डेटा के लिए उपयोग किया जाता है और उन डेटा के लिए भी होता है जिनमें चरम मान होते हैं। यह डेटा का मध्य मूल्य है जब डेटा को आरोही या अवरोही क्रम में व्यवस्थित किया जाता है। डेटा मूल्यों

की सम संख्या के मामले में दो मध्य मूल्य होंगे और हम औसत के रूप में उनका औसत लेते हैं। यह बड़े पैमाने पर जीवन-परीक्षण और अस्तित्व विश्लेषण में उपयोग किया जाता है। माध्यका (मीडियन) की प्रवित्ति होती है कि 50% डेटा मान के नीचे होगा और अन्य 50% औसत से ऊपर होगा। हम इसे 50वाँ प्रतिशत भी कहते हैं।

पहली चतुर्थक (Q1) एक मान है जिसके नीचे लगभग 25% डेटा है। तीसरा चतुर्थक (Q3) एक मान है जिसके नीचे लगभग 75% डेटा है। इस नियम से Q2 मंझला होगा।

8.6.3 बहुलांक (मोड)

वह मान जो किसी डेटा में अधिकतम संख्या में होता है, बहुलांक (मोड) कहलाता है। एकल मोड, दो मोड और कभी-कभी कई मोड हो सकते हैं। यदि कोई मान दोहराया नहीं गया है, तो कोई मोड नहीं है उदाहरण यदि डेटा {2, 5, 8, 1, 4, 9, 6} है।

8.7 विचलनशीलता के माप

विचलनशीलता मापे गए परिमाण (वैल्यू) की एक अंतर्निहित विशेषता है। यह कई कारणों से होता है जिनमें से कुछ को नियंत्रित नहीं किया जा सकता है। किसी लक्ष्य के आसपास डेटा मानों का प्रसार फैलाव या बिखराव कहलाता है। एक स्थिर या सुसंगत डेटा में अस्थिर डेटा की तुलना में कम फैलाव होगा। विचलन के संख्यात्मक मापों को प्रकीर्णन (फैलाव) के माप या प्रसार के माप *(dispersion measures or measures of spread)* कहा जाता है।

कुछ माप निम्नलिखित हैं।

8.7.1 प्रसार (रेंज)

यह डेटा के सबसे बड़े और सबसे छोटे परिमाणों के बीच का अंतर होता है। यह तब उपयोगी होता है जब डेटा वयस्क पुरुष की ऊंचाई की तरह काफी स्थिर होता है। जब 'प्रसार' अधिक होता है, तो इसका मतलब है कि डेटा में उच्च भिन्नता है। यह स्थिति तब होती है जब डेटा में कुछ असामान्य मान होते हैं। कभी-कभी 'प्रसार' को {न्यूनतम अधिकतम} के रूप में निर्दिष्ट किया जाता है, लेकिन दो या अधिक डेटा सेटों की तुलना करने के लिए हमें प्रसार (रेंज)=(अधिकतम-न्यूनतम) का उपयोग करना पड़ता है, जो एक एकल मूल्य देता है। 0.6 मिलीग्राम/एल फ्लोराइड स्तर का एक प्रसार 1.1 मिलीग्राम/एल के प्रसार से कम भिन्नता को इंगित करता है।

8.7.2 विचरण और प्रमाप विचलन

माध्य (M) के आस-पास डेटा मानों के प्रसार का विचरण कहते हैं। यदि कई मान, माध्य से दूर हैं, तो हम उच्च विचरण कहते हैं और यदि कई मान, माध्य के करीब हैं, तो हमें कम विचरण मिलता है। जनसंख्या के विचरण को $\sigma 2$ द्वारा निरूपित किया जाता है और इसका सूत्र होता है

$$\sigma^2 = \frac{\sum (X_i - M)^2}{N}$$

जहाँ N जनसंख्या में इकाइयों की संख्या है।

यह हमेशा सकारात्मक होता है, लेकिन वर्ग (squared) इकाइयों में व्यक्त किया जाता है। उदाहरण के लिए, यदि ऊँचाई को सेंटीमीटर में मापा जाता है, तो विचरण को ''सेंटीमीटर वर्ग' में व्यक्त किया जाना चाहिए, जो कि समझ में आना मुश्किल है।

प्राकृतिक इकाइयों में भिन्नता को मापने के लिए हम मानक विचलन (SD) का उपयोग करते हैं जो कि दिए गए विचरण का सकारात्मक वर्ग मूल है

$$s = \sqrt{\frac{\sum (x_i - M)^2}{N}}$$

निम्नलिखित नमूने का उपयोग करके मानक विचलन (s) को n 'प्रतिदर्श का आकार' से गणना की गई यह नमूना मानक विचलन σ का अनुमान है।

$$s = \sqrt{\frac{\sum (X_i - \overline{X})^2}{n}}$$

जहाँ $\overline{X}$ का मतलब प्रतिदर्श का माध्य है। छोटे नमूनों के मामले में, हम दिए गए मानक विचलन (s) के लिए एक अलग सूत्र का उपयोग करते हैं

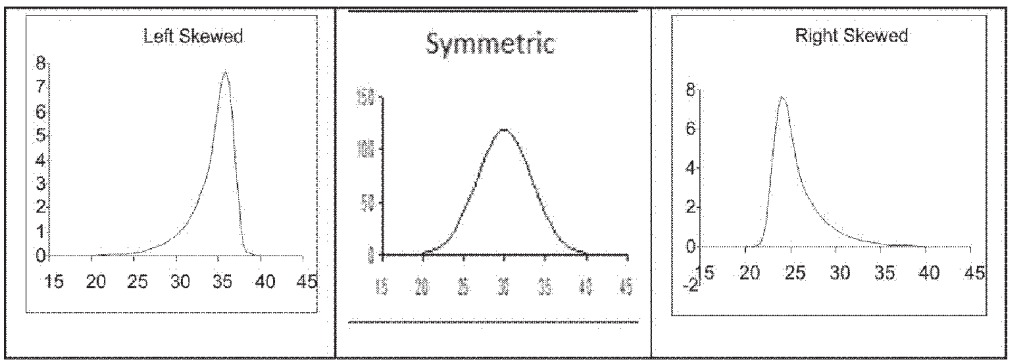
$$s = \sqrt{\frac{\sum (X_i - \overline{X})^2}{(n-1)}}$$

नोटः विभाजक में (n-1) है और यह छोटे और बड़े दोनों नमूनों के लिए उपयोगी है।

विचरण के दो अन्य माप हैं जो डेटा वितरण (या हिस्टोग्राम) के आकार का वर्णन करने के लिए उपयोग किए जाते हैं। इन्हें नीचे उल्लिखित किया गया है।

8.7.3 वैषम्य (Skewness)

यह वितरण में समरूपता की कमी का एक माप है। जब वितरण में केंद्रीय मान (माध्य) के नीचे और ऊपर समान मान होते हैं तब हम कहते हैं कि वितरण सममित है। लंबी-बाईं तरफ के झुकाव वाले वितरण को *बाएं-वैषम्य* कहा जाता है। उसी तर्क के साथ, एक *दाहिने-वैषम्य* वितरण में दाएं तरफ झुकाव होगा। ऐसे वितरण निचे चित्र 8.2 में दिखाए गए हैं।



चित्र 8.2: पैटर्न (आकार) सममित और असममित वितरण

कार्ल पियर्सन के गुणांक का उपयोग करके वैषम्य (तिरछेपन) को मापा जाता है

$$S_k = \frac{3(Mean - Median)}{SD}$$

और यह मान धनात्मक, ऋणात्मक या शून्य हो सकता है। यदि वितरण सममित रूप से होता है तो उस स्थिति में $S_k=0$ क्योंकि माध्य = माध्यिका होता है।

8.7.4 कुकुदता (Kurtosis)

यह वितरण के आकार में चरमता का एक माप है। कुछ वितरण ऊंचे शिखर पर होते हैं जबकि कुछ सपाट दिखते हैं। वितरण जो न तो बहुत लंबा है और न ही बहुत सपाट है सामान्य के रूप में जाना जाता है। एक वितरण जो सामान्य से अधिक ऊंचा होता है, उसे लेप्टोकर्टिक वितरण के रूप में जाना जाता है और जो सामान्य से कम शिखर पर होता है, उसे प्लैटीकर्टिक के रूप में जाना जाता है। एक अच्छी तरह से व्यवहार किए गए डेटा के लिए कुकुदता (कर्टोसिस) का मूल्य होगा जीव विज्ञान और स्वास्थ्य विज्ञान में बुनियादी सांख्यिकी के कई विवरण सुंदर राव और रिचर्ड (2012) में पाए जा सकते हैं।

निम्नलिखित अनुभाग में हम सांख्यिकीय अनुमान के कुछ बुनियादी विचारों को सीखेंगे।

8.8 सार्थकता परिक्षण

सांख्यिकीय तरीके न केवल डेटा पैटर्न का वर्णन करने और डेटा को सारांशित करने में मदद करते हैं, बल्कि प्रतिदर्श डेटा के आधार पर किसी जनसंख्या की अज्ञात विशेषताओं के बारे में सार्थक निष्कर्ष निकालने में भी मदद करते हैं। आंकड़ों के इस क्षेत्र को आनुमानिक सांख्यिकी के रूप में जाना जाता है। इनका केन्द्र-बिंदु दो महत्वपूर्ण क्षेत्रों में होता है;

क) अज्ञात मापदंडों का अनुमान लगाना (जैसे 'कम जन्म के वजन की व्यापकता')

ख) प्रतिदर्श डेटा का उपयोग करके जनसंख्या विशेषताओं के बारे में एक परिकल्पना (विश्वास) की सच्चाई का परीक्षण करना।

परिकल्पना के सांख्यिकीय परीक्षण, सांख्यिकीय अनुमान का एक हिस्सा हैं। एक परिकल्पना एक जनसंख्या के अज्ञात मापदंडों के बारे में एक संख्यात्मक कथन है। शाब्दिक अर्थों में एक परिकल्पना शोधकर्ता द्वारा किया गया विश्वास या एक सत्यापित कथन है। संक्षेप में, हम यह जांचना चाहते हैं कि प्रतिदर्श पर आधारित निष्कर्ष केवल संयोग मात्र हैं (सार्थक अंतर नहीं) या क्या उन्हें ज्ञात कारकों के लिए जिम्मेदार ठहराया जा सकता है? परिकल्पना के महत्व और परीक्षण की अवधारणा के बीच एक सूक्ष्म अंतर है लेकिन सभी लोग उसका अनुप्रयोग करते वक्त उसे एक सामान ही व्यक्त करते हैं।

एक या एक से अधिक परिकल्पनाओं को अध्ययन करने से पहले उनका निरूपण किया जाता है और इन परिकल्पनाओं की सच्चाई को प्रतिदर्श डेटा के प्रकाश में सत्यापित किया जाता है, जो सबूत के रूप में, शोधकर्ता द्वारा एकत्र किया गया हो। स्पष्ट गणितीय कारणों से हम एक परिकल्पना के साथ शुरू करते हैं कि कोई प्रभाव या कोई घटना नहीं होगी और यह जांच करते हैं कि इसके सच होने की कितनी संभावना है। जवाब केवल संभावना के संदर्भ में दिखाई देता है।

हमें निम्नलिखित तकनीकी शब्दों को जानना होगा।

1) शून्य परिकल्पनाः H^O द्वारा निरूपित यह एक कथन है जिसमें एक शून्य प्रभाव या प्रभाव की अनुपस्थिति का उल्लेख होता है। इसे रुचि के मानदंडों को संबोधित करने वाले एकल माप के रूप में माना जाता है। कभी-कभी इसे बिना किसी अंतर वाली परिकल्पना के रूप में भी माना जाता है। यहां कुछ उदाहरण दिए गए हैं

क) H^O: प्रशिक्षण से पहले और बाद का औसत ज्ञान स्कोर समान रहता है।

ख) H^O: अध्ययन क्षेत्र में बच्चों में कोई स्टंटिंग (वृद्धि की कमी) नहीं है।

ग) H^O: किसी दिए गए गाँव में धूम्रपान करने का प्रचलन 30% है।

2) वैकल्पिक परिकल्पनाः जब शून्य परिकल्पना डेटा द्वारा समर्थित नहीं होती है, तो हम कहते हैं कि इसे अस्वीकार कर दिया गया है और हम H1 द्वारा निरूपित वैकल्पिक परिकल्पना नामक एक अन्य कथन को स्वीकार करने के लिए सहमत हैं। या तो शून्य या वैकल्पिक परिकल्पना सही होगी लेकिन किसी दिए गए संदर्भ में दोनों नहीं।

नीचे दिए गए वैकल्पिक परिकल्पना को निर्दिष्ट करने के दो तरीके हैं।

क) एक तरफा विकल्पः इस पद्धति में हम परिणाम की दिशा निर्दिष्ट करते हैं, अगर यह शून्य नहीं है। उदाहरण के लिए, धूम्रपान का प्रचलन <30% 'एक तरफा विकल्प है। हम विकल्प के रूप में > 30% 'धूम्रपान की व्यापकता पर भी विचार कर सकते हैं।

ख) दो-तरफा विकल्पः इस पद्धति में हम परिणाम की दिशा निर्दिष्ट नहीं करते हैं। यह सकारात्मक या नकारात्मक हो सकता है। उदाहरण के लिए, '30% के बराबर धूम्रपान का प्रचलन' दो तरफा वैकल्पिक परिकल्पना है।

3) टाइप- I और टाइप- II त्रुटियांः चूंकि अशक्त परिकल्पना (HO) पर निर्णय एक प्रतिदर्श पर आधारित है, इसलिए जनसंख्या पर हम HO को अस्वीकार करने की संभावना रखते हैं भले ही यह वास्तव में सही था (प्रतिदर्श खराब हो सकता है)। यह एक गलत अस्वीकृति है और टाइप-I त्रुटि कहा जाता है। इसी तरह, हम गलत स्वीकृति करके टाइप - II त्रुटि कर सकते हैं। इन दोनों त्रुटियों से पूरी तरह से बचा नहीं जा सकता है, लेकिन हम त्रुटि दरों को ठीक कर सकते हैं और एक सांख्यिकीय प्रक्रिया विकसित कर सकते हैं।

4) सार्थकता का स्तरः झूठी अस्वीकृति की अधिकतम सहनीय दर को ग्रीक अक्षर α (अल्फा) द्वारा निरूपित किया जाता है। आमतौर पर इसका मान 5% के रूप में लिया जाता है, हालांकि हम कभी-कभी 1% लेते हैं। हम $\alpha = 0.05$ लिखते हैं इसका मतलब है कि 5% उदाहरणों में प्रक्रिया H0 को अस्वीकार कर सकती है, भले ही यह वास्तव में सच हो।

5) क्रांतिक मान (क्रिटिकल वैल्यू): परीक्षण प्रक्रिया में, सूत्र का उपयोग करने के बाद, सैंपल डेटा से प्राप्त मान, परीक्षण मान कहलाता है। इस मान की तुलना एक क्रांतिक मान या सीमा मान के साथ की जाती है जो सांख्यिकीय तालिकाओं में उपलब्ध होता है। ये क्रांतिक मान परीक्षण के प्रकार और α के मान पर आधारित होता है। जब परीक्षण मान, क्रांतिक मान से अधिक हो जाता है, तो हम H0 को 5% सार्थकता के स्तर पर अस्वीकार कर देते हैं। इसका मतलब है कि शून्य परिकल्पना के सच होने की संभावना बहुत कम है। बड़े नमूनों (30 से कम नहीं) पर आधारित परीक्षणों के लिए, दो तरफा विकल्प के लिए 5% के स्तर पर क्रांतिक मान 1.96 है और एक तरफा विकल्प के लिए, क्रांतिक मान 1.65 है।

6) परीक्षण की शक्तिः यह संभावना भी होती है की हम वैकल्पिक परिकल्पना को स्वीकार करने में सक्षम हैं जब यह वास्तव में सच है। इसे (1-β) द्वारा निरूपित किया जाता है जहां β झूठी स्वीकृति की दर को दर्शाता है। इस प्रकार, परीक्षण

की शक्ति, सच्ची स्वीकृति की दर है। आमतौर पर, शोधकर्ता 80% परीक्षण की शक्ति या अधिक के साथ परीक्षणों को देखते हैं। इसका अर्थ है β= 0-20।

7) एक परीक्षण का पी-मानः यह तालिकाओं से क्रांतिक मान का उपयोग करने के लिए एक वैकल्पिक दृष्टिकोण है। पी-मान प्रतिदर्श डेटा के आधार पर टाइप-I त्रुटि की वास्तविक संभावना है। इसकी तुलना α से की जाती है। यदि पी-मान α से कम है, तो हम शून्य परिकल्पना को अस्वीकार करते हैं और कहते हैं कि निष्कर्ष महत्वपूर्ण हैं। एक नियम के रूप में, छोटे पी-मूल्य सांख्यिकीय महत्व दर्शाता है। यदि पी-मान α से अधिक है, तो हम कहते हैं कि 'निष्कर्ष भी संयोग के कारण हो सकते हैं'। पी-मूल्य की गणना कंप्यूटर गहन प्रक्रिया है।

अपनी प्रगति जांचें

4) यादृच्छिक निदर्शन की विभिन्न विधियों का विश्लेषण कीजिए?

.....

.....

.....

.....

.....

5) केन्द्रीय प्रविति के माप क्या हैं ? अंकगणितीय माध्य की व्याख्या कीजिए?

.....

.....

.....

.....

.....

6) टिपण्णी लिखिए क) शून्य परिकल्पना, ख) पी-मान ।

.....

.....

.....

.....

.....

निम्नलिखित खंड में हम सार्वजनिक स्वास्थ्य अध्ययन में आमतौर पर उपयोग किए जाने वाले कुछ सांख्यिकीय परीक्षणों पर चर्चा करेंगे।

8.8.1 अनुपात परीक्षण

इस परीक्षण का उद्देश्य किसी बीमारी के देखे गए प्रचलन की तुलना काल्पनिक दृष्टि से

करना है। उदाहरण के लिए, शोधकर्ता यह बताता है कि 250 व्यक्तियों के प्रतिदर्श का प्रचलन 35% है। यह काल्पनिक रूप से माना जाता है कि जनसंख्या में प्रचलन 50% है। हम परीक्षण करना चाहते हैं कि क्या प्रचलित के काल्पनिक और देखे गए मूल्यों के बीच अंतर महत्वपूर्ण है। इसे एक नमूना के लिए अनुपात परीक्षण कहा जाता है क्योंकि प्रतिशत और अनुपात समान अर्थ को व्यक्त करते हैं।

अनुपात के लिए एक नमूना परीक्षणः H^0: p = po (अनुपात के रूप में व्यक्त काल्पनिक मान) और H1: $p \neq p_0$ (दो तरफा विकल्प)। सूत्र का उपयोग करके परीक्षण मान की गणना की जाती है

$$Z = \frac{p - p_o}{\sqrt{\frac{P_o(1-p_o)}{n}}}$$

गणक, अंतर है और भाजक को मानक त्रुटि कहा जाता है। यह मान या तो सकारात्मक या नकारात्मक हो सकता है लेकिन हम संकेत को अनदेखा करते हैं और परीक्षण मूल्य को पढ़ते हैं। α=0.05 लेना क्रांतिक मान 1.96 है। यदि Z> 1-96 HO को अस्वीकार करता है और यह निष्कर्ष निकालता है कि अंतर महत्वपूर्ण है (संयोग से घटना नहीं)।

उदाहरण 8.4: 150 स्कूली बच्चों पर एक प्रतिदर्श अध्ययन में, यह पाया गया कि 26 छात्रों को सुनने में कठिनाई हुई। क्या यह अध्ययन इस कथन (विश्वास) का समर्थन करेगा कि सामान्य रूप से 20% स्कूली बच्चों को सुनने में कठिनाई होती है?

समाधानः यहाँ n = 150 और नमूना अनुपात p=26/150 = 0.17 या 17% है। शून्य परिकल्पना HO: p = 0.20 (p0) और H1: p = 20. 0.20 (दो विकल्प वाली परिकल्पना) है। अब इसका पता लगाते हैं

अ) अंतर = 0.17-0.20 = -0.03

ब) स्टैंडर्ड एरर = $\sqrt{\frac{P_0(1-p_0)}{n}} = \sqrt{\frac{0.20*0.80}{150}} = 0.033$

स) टेस्ट वैल्यू (Z) = 0.03 / 0.033 = 0.909

α = 0.05, क्रांतिक मान 1.65 है और Z> 1-65. तब हम शून्य परिकल्पना को स्वीकार नहीं कर सकते हैं और इसलिए अंतर महत्वपूर्ण है। हम शोधकर्ता के इस विश्वास को स्वीकार कर सकते हैं।

अनुपात के लिए दो प्रतिदर्श परीक्षणः हम यह परीक्षण करना चाहते हैं कि क्या दो स्वतंत्र समूहों से प्राप्त अनुपात के बीच का अंतर सांख्यिकीय रूप से महत्वपूर्ण है। यदि P1 और P2 दो अनुपातों को निरूपित करते हैं, तो हम शून्य परिकल्पना को H0: P1 = P2 (अंतर शून्य) और H1: P1 = P2 (दो तरफा विकल्प) के रूप में फ्रेम करते हैं। परीक्षण मान के रूप में गणना की जाती है.

$$Z = \frac{p_1 - p_2}{\sqrt{\frac{pq}{n_1} + \frac{pq}{n_2}}}$$

जहाँ $p = \frac{n_1 p_1 + n_2 p_2 .}{n_1 + n_2}$ संयुक्त अनुपात कहलाता है और q = (1-p) | α = 0.05 ने से हमें 1.96 के रूप में क्रांतिक मान मिलता है। यदि Z> 1.96 HO को अस्वीकार करता है और विचार करता है कि अंतर महत्वपूर्ण है।

उदाहरण 8.5: एक सामुदायिक स्वास्थ्य अध्ययन में एक कोहोर्ट ए को कवर करते हुए

यह पाया गया है कि 120 में से 22 गर्भवती महिलाएं एनीमिक हैं और कोहॉर्ट बी में 150 गर्भवती महिलाओं में से 41 एनीमिक पाई जाती हैं। 5% के महत्व के स्तर पर, क्या हम इस बात पर विचार कर सकते हैं कि कोहर्ट बी में कोहोर्ट ए की तुलना में अधिक एनीमिक महिलाएं हैं?

समाधानः यहाँ n1 = 120 और n2 = 150. नमूना अनुपात P1 = 22/120 = 0.18 और P2 = 41/150 = 0.27 है। शून्य परिकल्पना H0: P1 = P2 (कोई अंतर नहीं) और H1: P1 <P2 (एक तरफा परिकल्पना) है। अब इसका पता लगाते हैं

ए) अंतर = 0.18-0.27 = -0.09

बी) संयुक्त अनुपात $(p) = \frac{120*0.18+150*2.27}{120+150} = 0.17$

सी) q = 1-0.17 = 0.83

डी) मानक त्रुटि = $\sqrt{\left\{\frac{0.17*0.83}{120}+\frac{0.17*0.83}{150}\right\}} = \sqrt{0.0012+0.0009} = 0.046$

ई) परीक्षण मान (Z) = 0.09 / 0.046 = 1.96 (नकारात्मक संकेत की अनदेखी)

α= 0.05 लेना क्रांतिक मान 1.65 है। Z> 1.65 तब हम शून्य परिकल्पना को स्वीकार नहीं कर सकते हैं और इसलिए अंतर महत्वपूर्ण है। हम शोधकर्ता के इस विश्वास को स्वीकार कर सकते हैं कि कोहोर्ट बी में कॉहोर्ट ए की तुलना में अधिक एनीमिक महिलाएं हैं।

8.8.2 माध्य परीक्षण (टी-टेस्ट)

इन परीक्षणों के साथ हम एक काल्पनिक माध्य के साथ एक विशेषता के देखे गए प्रतिदर्श माध्य की तुलना कर सकते हैं। हम दो स्वतंत्र या आश्रित नमूनों के बीच अंतर के महत्व के लिए भी परीक्षण कर सकते हैं। यह माना जाता है कि व्यक्तिगत डेटा मान सामान्य वितरण का पालन करते हैं।

- जब नमूना का आकार बड़ा होता है और शून्य परिकल्पना सच होती है, तो परीक्षण मान सामान्य वितरण का अनुसरण करता है और परीक्षणों को Z- परीक्षण (R.A. फिशर द्वारा प्रस्तावित) के रूप में जाना जाता है।
- छोटे नमूनों के साथ परीक्षण मान की सामान्य धारणा अच्छी नहीं है। इस मामले में हम एक विशेष वितरण का उपयोग करते हैं जिसे स्टूडेंट टी-वितरण कहा जाता है (डल्यू. एस. गोस्सेट द्वारा प्रस्तावित जिसका कल्पित नाम स्टूडेंट था) है। इन परीक्षणों को टी-परीक्षण के रूप में जाना जाता है।
- दिलचस्प बात यह है कि टी-टेस्ट को छोटे और बड़े दोनों नमूनों के लिए लागू किया जा सकता है जबकि जेड-परीक्षण को छोटे नमूनों पर लागू नहीं किया जा सकता है।

दो प्रतिदर्श माध्य के टी-टेस्टः यह दो स्वतंत्र प्रतिदर्शों में देखे गए विशेषता के माध्य के बीच अंतर की तुलना करने के लिए एक परीक्षण है। मान लीजिए कि हीमोग्लोबिन को रोगियों के दो स्वतंत्र नमूनों से मापा जाता है। एक समूह को उच्च प्रोटीन आहार और दूसरे को सामान्य आहार के साथ इलाज किया जाता है। हम परीक्षण करना चाहते हैं कि क्या प्रतिदर्श के अंतर महत्वपूर्ण है।

शून्य परिकल्पना HO हैः अंतर शून्य है। दो तरफा विकल्प को H1 के रूप में लिया जा सकता हैरू अंतर शून्य नहीं है। मान लें कि दो समूहों के लिए, नमूना आकार n1 और n2 हैं और माध्य क्रमशः और हैं। इसके अलावा s1 और s2 क्रमशः दो समूहों में मूल्यों के मानक विचलन हैं।

हमें संयुक्त द्वारा चिह्नित किए गए मानक विचलन को खोजना होगा

$$S = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

परीक्षण मान के रूप में गणना की जाती है

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

क्रांतिक मान, स्वतंत्रता के अंश के साथ टी-वितरण की तालिकाओं में पाया जाता है। यदि टी की गणना का मान, क्रांतिक मान से अधिक है, तो हम अंतर को महत्वपूर्ण मानते हैं।

कभी-कभी हमें एक ऐसी स्थिति मिलती है जहां एक हस्तक्षेप (जैसे उपचार, प्रशिक्षण आदि) देने से पहले और बाद में प्रत्येक विषय के लिए एक माप लिया जाता है। ऐसे मामलों में हमें युग्मित टी-टेस्ट का उपयोग करना होगा। इस परीक्षण में द मानों का केवल एक ही नमूना होता है, लेकिन प्रत्येक मामले के लिए उपचार के बाद एक और पहले दो अवलोकन होंगे। इस तरह के डेटा को कभी-कभी 'पूर्व-पोस्ट डेटा' कहा जाता है। हम पूर्व और बाद के माध्य के बीच अंतर के महत्व का परीक्षण करना चाहते हैं।

हम इस इकाई को एक नोट के साथ समाप्त करते हैं कि आँकड़े अनिश्चितता की उपस्थिति में वास्तविक जीवन की घटनाओं को समझने के लिए विज्ञान है और सत्य की खोज विश्लेषण का लक्ष्य है।

8.9 सारांश

इस पाठ में हमने निम्नलिखित बातें सीखी हैं।

सांख्यिकी डेटा संग्रह और सार्वजनिक स्वास्थ्य अध्ययन के विश्लेषण में मदद करता है। आँकड़ों का महत्व सारणी, चार्ट और औसत जैसे माध्य में डेटा के विशाल मात्रा को संक्षेप में प्रस्तुत करने में निहित होती है।

निदर्शन डेटा संग्रह की आदर्श विधि है क्योंकि जनसंख्या का कुल सर्वेक्षण कठिन, महंगा और समय लेने वाला है। चयन में अन्वेषक के पूर्वाग्रह से बचने के लिए सरल यादृच्छिक प्रतिदर्श, स्तरीकृत निदर्शन और क्लस्टर निदर्शन जैसी कई नमूने विधियां उपलब्ध हैं।

सांख्यिकीय डेटा का अर्थ, भिन्नता और मानक विचलन जैसे सारांश मूल्यों की सहायता से किया जाता है। चूंकि, सभी नमूना मूल्य जनसंख्या सुविधाओं (मापदंडों) के अनुमान हैं, वे निदर्शन त्रुटियों के अधीन हैं। इस कारण से हम यह सुनिश्चित करने के लिए परिकल्पना का परीक्षण करते हैं कि इनमें त्रुटियां जो वादा की गई हैं उससे अधिक नहीं हैं (परीक्षण का महत्व और शक्ति का स्तर) ।

8.10 संदर्भ

इंद्रायन, ए. एंड सत्यनारायना,एल. (2006). *बायोस्टटिस्टिक्स फॉर मेडिकल, नर्सिंग एंड फार्मेसी स्टूडेंट्स.* न्यू दिल्लीः प्रेंटिस हॉल ऑफ इंडिया.

राव, पी.एस., एंड रिचर्ड, जे. (2012). *इंट्रोडक्शन टू बायोस्टैटिस्टिक्स एंड रिसर्च मैथड्स.* 5वां संस्करण. नई दिल्लीः प्रेंटिस हॉल ऑफ इंडिया.

थायिल, जे. एंड जीजा, एम. सी. (2013). इश्यू ऑफ क्रियेटिंग ए न्यू कैड्री ऑफ डॉक्टर्स फॉर रूरल इंडिया. *इंटरनेशनल जर्नल ऑफ मेडिसिन एंड पब्लिक हेल्थ,* 3: 8-11.

8.11 आपकी प्रगति की जांच करने के लिए उत्तर

1) सांख्यिकी प्रतिदर्श डेटा के आधार पर जनसंख्या के बारे में निष्कर्ष निकालने में मदद करती है। विवरण के लिए भाग 8.0 देखें।

2) एक जनसंख्या अध्ययन के लक्ष्य से संबंधित सभी विषयों (लोगों, जानवरों, पौधों आदि) का संग्रह है और इसे कोहोर्ट या अध्ययन समूह के रूप में भी जाना जाता है। एक प्रतिदर्श जनसंख्या का एक प्रतिनिधि भाग (उप सेट) है। विवरण के लिए भाग 8.1 देखें।

3) एक प्रतिदर्श प्रारूप एक योजना है जिसके अनुसार अध्ययन में डेटा का संकलन किया जाएगा। यह निम्नलिखित पहलुओं को निर्दिष्ट करता है। (ए) नमूना फ्रेम (बी) नमूना आकार (सी) नमूनाकरण की विधि (डी) प्रश्नावली का डिजाइन (ई) डेटा प्रविष्टि और सत्यापन (एफ) डेटा के लिए विश्वसनीयता के उपाय।

4) यादृच्छिक निदर्शन में, जनसंख्या के सदस्यों (इकाइयों) को यादृच्छिक या लॉटरी प्रकार के तंत्र द्वारा प्रतिदर्श में शामिल किया जाता है। रैंडम सैंपलिंग के विभिन्न तरीके हैं: (क) सिंपल रैंडम सैंपलिंग (ख) सिस्टेमेटिक रैंडम सैंपलिंग (ग) स्तरीकृत रैंडम सैम्पलिंग (घ) क्लस्टर सैम्पलिंग मल्टी स्टेज सैम्पलिंग। विवरण के लिए भाग 8.2 देखें।

5) केंद्रीय प्रवृत्ति के विभिन्न माप हैं: (ए) मीन (बी) मेडियन (सी) मोड। डेटा का सबसे अधिक इस्तेमाल किया जाने वाला औसत अंकगणित माध्य या केवल माध्य है। यह बस मूल्यों की संख्या से विभाजित सभी मूल्यों का योग है। विवरण के लिए भाग 8.6 देखें।

6) शून्य परिकल्पना एक कथन है जो एक अशक्त प्रभाव या प्रभाव की अनुपस्थिति का उल्लेख करता है। इसे HO द्वारा दर्शाया गया है।

तालिकाओं से क्रांतिक मानों का उपयोग करने के लिए पी-मान एक वैकल्पिक दृष्टिकोण है। पी-मान प्रतिदर्श डेटा पर आधारित टाइप I त्रुटि की गणना की वास्तविक संभावना है। विवरण के लिए भाग 8.8 देखें।

इकाई 9 डेटा विश्लेषण*

इकाई की रूपरेखा

अध्ययन के उद्देश्य

इस इकाई को पढ़ने के बाद, आप निम्न कार्य कर सकेंगेः

- एक्सेल में एक डेटा फाइल बनाना;
- एक्सेल शीट में डेटा संपादित करना;
- अन्य सॉफ्टवेयर में डेटा को सुरक्षित करना, पुनः प्राप्त करना और निर्यात करना;
- एक्सेल के साथ बुनियादी सांख्यिकीय विश्लेषण करना;
- एक्सेल डेटा का उपयोग करके SPSS के साथ काम करना; तथा
- SPSS के साथ प्रतिगमन विश्लेषण और एनोवा करना।

9.0 परिचय

माइक्रोसॉफ्ट (MS) एक्सेल एक स्प्रेडशीट पैकेज है जिसे सरल डेटा सेट को व्यवस्थित करने, सामान्य गणना करने, ग्राफ बनाने और सांख्यिकीय विश्लेषण को संभालने के लिए डिजाइन किया गया है। यह एमएस-ऑफिस समूह का एक भाग है जिसमें वर्ड, एक्सेल, पावर पॉइंट और एक्सेस होते हैं। एक्सेल कई अंतर्निहित प्रोग्राम प्रदान करता है, जो सरल 'माउस' संचालन के साथ चलते हैं।

हम रोगी के डेटा को संग्रहीत करने के लिए एक्सेल का उपयोग कर सकते हैं, (केस) प्रकरण दर केस के आधार पर बड़ी संख्या में विभिन्न मापदंडों से संबंधित डेटा को संग्रहित कर सकते हैं। हम किसी भी संख्या के डेटा पर कुल, अंतर, प्रतिशत और यहां तक कि वैज्ञानिक गणना जैसी सरल गणना कर सकते हैं। कई संगठन डेटा साझा करने के लिए एक्सेल का उपयोग प्लेटफॉर्म के रूप में करते हैं। कई वेब साइट एक्सेल प्रारूप में डेटा

* योगदानकर्ता – प्रो. के.वी.एस. सरमा (सेवानिवृत्त), सांख्यिकी विभाग, श्री वेंकटेश्वर विश्वविद्यालय, तिरुपति.
अनुवादक – डॉ. निशीथ राय, सहायक प्रोफेसर, मानवविज्ञान, म.गां.अ.हिं.वि. वर्धा, महाराष्ट्र।

डाउनलोड करने का विकल्प प्रदान करती हैं (जैसे आपके बैंक खाते का विवरण)। एक्सेल स्मार्ट फोन और टैबलेट में भी उपलब्ध है (WPS) ऑफिस एक ऐसा ऐप है)।

सांख्यिकीय अनुप्रयोगों के लिए कई सॉफ्टवेयर पैकेज हैं, लेकिन उनमें से कई महंगे हैं। Excel बुनियादी सांख्यिकीय विश्लेषण करने के लिए कुछ 'टूल' प्रदान करता है।

आइए हम यह समझना शुरू करें कि एक्सेल में डेटा फाइल कैसे बनाई जा सकती है।

9.1 एक्सेल में एक डाटा फाइल बनाना

जब MS-Office कंप्यूटर में 'इंस्टाल' होता है, तो हम एक्सेल बटन को टास्क बार या स्टार्ट मेन्यू में पाते हैं। इस बटन पर क्लिक करने से एक्सेल खुल जाता है। Book1 नाम की एक नई कार्यपुस्तिका खुलेगी और इसमें आम तौर पर शीट 1, शीट 2 और शीट 3 नाम की शीट शामिल होती है और सक्रिय शीट शीट 1 है। एक्सेल में डेटा फाइल को वर्कबुक कहा जाता है। हम चर्चा के लिए एक्सेल 2010 का उल्लेख करते हैं।

एक्सेल वर्कशीट की कुछ विशेषताएं निम्नलिखित हैं।

क) हर शीट में A, B,… Z, AA, AB,… XFD जैसे 16384 कॉलम होते हैं।

ख) एक शीट में पंक्तियों की संख्या 1048576 है, जिसे 1, 2, 3 के रूप में गिना जाता है।

ग) शीट के प्रत्येक सेल की पहचान उसके सेल एड्रेस से की जाती है। उदाहरण के लिए, A3 का अर्थ है स्तंभ A और पंक्ति 3 के प्रतिच्छेदन पर कोशिका (सेल)।

घ) प्रत्येक सेल का उपयोग विभिन्न प्रकार के डेटा जैसे संख्याओं, पाठ, दिनांक, समय, मुद्रा आदि को दर्ज करने के लिए किया जाता है। डिफॉल्ट स्थिति को सामान्य कहा जाता है, जिसके लिए कोई विशिष्ट प्रारूप निर्दिष्ट नहीं किया जाता है।

च) पहली पंक्ति आमतौर पर डेटा फाइल के कॉलम हेडिंग को इंगित करने के लिए उपयोग की जाती है। यदि विश्लेषण के लिए शीट का अर्थ है, तो कई शीर्षक नहीं होने चाहिए। हर कॉलम में केवल एक शीर्षक (हेडिंग) होना चाहिए।

टिप्पणीः कुछ उपयोगकर्ता बीच-बीच में एक उप शीर्ष प्रदान करके डेटा सेट को एक के बाद एक ऊपर से नीचे टाइप करते रहते हैं। यह सही नहीं है।

निम्नलिखित उदाहरण पर विचार करें।

उदाहरण 9.1: एन.टी.आर. स्वास्थ्य सेवा योजना 2016-17 के तहत आयोजित स्वास्थ्य शिविरों की संख्या पर एक विशिष्ट डेटा फाइल तालिका 9.1 में दर्शाई गई है। (स्रोतः सामाजिक आर्थिक सर्वेक्षण 2016-17, आंध्र प्रदेश सरकार)। आइए हम एक्सेल में एक डेटा फाइल बनाएँ।

प्रक्रियाः यहाँ पहला चरण है।

1) कीबोर्ड का उपयोग करके कोशिकाओं में डेटा दर्ज किया जाएगा। सेल में हर प्रविष्टि के बाद Enter बटन को दबाया जाना चाहिए।

2) 'टूलबार' में दिखाई देने वाले बाएँ, दाएँ और केंद्र संरेखण बटन का उपयोग करके डेटा को ठीक से संरेखित किया जा सकता है।

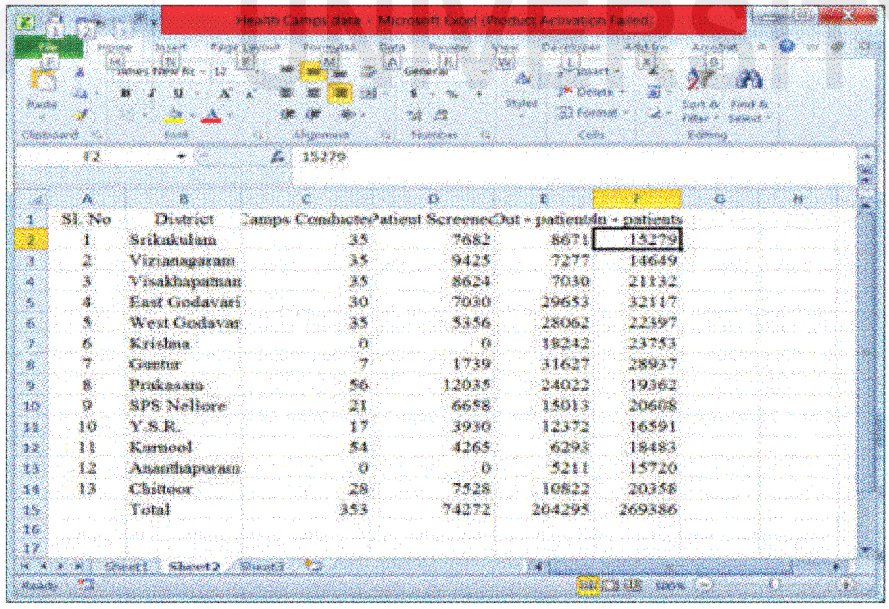
तालिका 9.1: स्वास्थ्य शिविर के आंकड़े (2016-17)

क्रम संख्या	जिला	आयोजित शिवरों की संख्या	मरीजों की संख्या	बाहरी मरीज (ओपीडी)	भर्ती मरीज
1	Srikakulam	35	7682	8671	15279
2	Vizianagaram	35	9425	7277	14649
3	Visakhapatnam	35	8624	7030	21132
4	East Godavari	30	7030	29653	32117
5	West Godavari	35	5356	28062	22397
6	Krishna	0	0	18242	23753
7	Guntur	7	1739	31627	28937
8	Prakasam	56	12035	24022	19362
9	SPS Nellore	21	6658	15013	20608
10	Y.S.R.	17	3930	12372	16591
11	Kurnool	54	4265	6293	18483
12	Ananthapuramu	0	0	5211	15720
13	Chittoor	28	7528	10822	20358
	Total कुल	353	74272	204295	269386

डेटा दर्ज करने के बाद, फाइल मेनू के 'सेव एज' विकल्प का उपयोग करके इस फाइल को सुरक्षित किया जाना चाहिए।

3) इस फाइल को 'हेल्थ कैंप डेटा' के रूप में डेस्कटॉप पर सेव करें।

4) एक्सेल में बनाई गई फाइल अब चित्र 9.1 में दिखाए गए चित्र की तरह दिखाई देगी।



चित्र 9.1: हेल्थ कैंप डेटा के लिए एक्सेल शीट

हम एक ही वर्कबुक की अलग-अलग शीट्स में या अलग-अलग वर्कबुक में अलग-अलग डेटा सेट बना सकते हैं। हम शीट 3 के दाईं ओर उपलब्ध बटन पर क्लिक करके नई वर्क शीट डाल सकते हैं। हम एक वर्कबुक में अधिकतम 256 वर्कशीट सम्मिलित कर सकते हैं।

9.2 एक्सेल शीट की संपादन विशेषताएं

एक्सेल शीट में कई ऐसे फीचर्स (विशेषताएं) हैं जो डाटा एंट्री को बेहद सरल बनाते हैं। उनमें से कुछ नीचे दिए गए हैं।

क) डेटा चयनः डेटा के एक हिस्से को माउस के साथ चुना जा सकता है ताकि चयनित क्षेत्र में परिवर्तन को किया जा सके (जैसे अक्षर आकार या फॉन्ट को बदलना)।

ख) कॉलम चौड़ाईः एक कॉलम की चौड़ाई आमतौर पर डिफॉल्ट रूप से 8.43 अंक के रूप में निर्धारित की जाती है और इसे मानक कहा जाता है। जब स्तंभ की सामग्री 8 या 9 वर्णों से अधिक होती है, तो बाकी छिप जाता है और दिखाई नहीं देता है। हम चयनित क्षेत्र में किसी भी दो स्तंभों के बीच लाइन पर माउस और डबल क्लिक के साथ सभी आवश्यक स्तंभों का चयन कर सकते हैं।

ग) फ्रीज पैनः जब डेटा में 22 से अधिक पंक्तियां होती हैं तो जब हम नीचे स्क्रॉल करना शुरू करते हैं तो शीर्षक गायब हो जाते हैं। इस तरह की स्थिति तब उत्पन्न होती है जब कॉलम शीर्षक बहुत व्यापक होता है। इससे डाटा एडिट करने में परेशानी होती है। पहली पंक्ति (शीर्षक पंक्ति) और पहले कॉलम को हमेशा दिखाई देने के लिए, मुख्य मेनू से फ्रीज पैन विकल्प का उपयोग करें। एक नियम के रूप में, सेल B2 पर माउस रख कर फ्रीज पैन पर क्लिक करें। यह B2 की पंक्ति को डार्क या गाढ़ा कर करेगा और पहली पंक्ति और पहला कॉलम दिखाई देगा। यदि आवश्यक नहीं है, तो व्यू → अनफ्रीज पैन का उपयोग करें।

घ) सॉर्ट/फिल्टरः यह विकल्प एक कॉलम में डेटा को छांटने में मदद करता है। जब आप बढ़ते या घटते क्रम में सॉर्ट कर सकते हैं तो पंक्तियों में अन्य सभी डेटा तत्व स्वचालित रूप से हल हो जाते हैं। फिल्टर विकल्प विशेष स्थिति को संतुष्ट करने वाले रिकॉर्ड का चयन करने में मदद करता है जैसे लिंग = 'एम' या आयु < 35 साल। सॉर्ट/फिल्टर बटन मुख्य मेनू में दिखाई देता है ।

च) कट, कॉपी और पेस्टः सेल या एक कॉलम या एक पूरी पंक्ति के एक समूह को शीट में किसी अन्य स्थान पर या कट, कॉपी और पेस्ट संचालन का उपयोग करके एक अलग शीट में कॉपी किया जा सकता है। हम पेस्ट के लिए Ctrl +V और कॉपी के लिए शॉर्ट कट्स Ctrl+ C का उपयोग कर सकते हैं।

छ) पेस्ट स्पेशलः यह एक महत्वपूर्ण संपादन सुविधा है। हम डेटा के एक हिस्से की प्रतिलिपि बना सकते हैं और बैक-एंड फॉर्मूले को परेशान किए बिना इसे दूसरे स्थान पर चिपका सकते हैं। यह 'वैल्यू' विकल्प चुनकर किया जाता है।

ज) वर्ड से डेटा निर्यात करनाः एक्सेल में प्राप्त डेटा या परिणामों को वर्ड दस्तावेज या पावर पॉइंट प्रेजेंटेशन में कॉपी और पेस्ट जा सकता है। इसी तरह, वर्ड टेबल में बनाये गये डेटा को भी एक्सेल में कॉपी और पेस्ट किया जा सकता है।

कोई सूत्र लिखकर योग, उप-योग, प्रतिशत आदि नई प्रविष्टियां भी सृजित करना संभव है। हमें कोड लिखने की आवश्यकता नहीं है। इसके बजाय हम उन सेल्स पर क्लिक कर सकते हैं जो सूत्र में शामिल हैं। एक्सेल की एक दिलचस्प विशेषता यह है कि परिणाम (जैसे योग) खोजने के बाद, यदि हम मूल डेटा में कोई परिवर्तन करते हैं, तो परिणाम स्वचालित रूप से अपडेट हो जाएंगे।

अपनी प्रगति जांचें

1) एक्सेल वर्कबुक की संरचना के बारे में लिखिए? एक शीट में कितनी 'सेल' पाई जाती हैं?

.....

.....

.....

2) एक्सेल शीट के किसी भी चार महत्वपूर्ण संपादन विशेषताओं का उल्लेख करें।

.....

.....

.....

3) एक्सेल में 'फ्रीज पैन' 'विकल्पों के बारे में संक्षेप में लिखें।

.....

.....

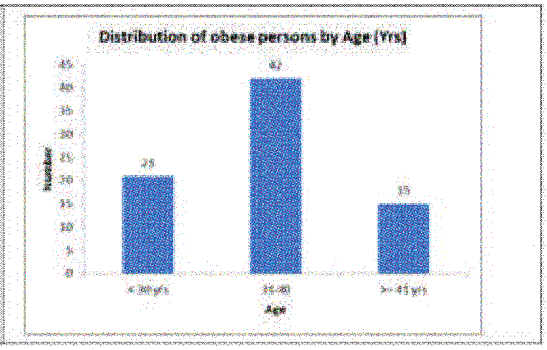
.....

9.3 एक्सेल में ग्राफ बनाना

एक्सेल के साथ विभिन्न प्रकार के सांख्यिकीय ग्राफ बनाए जा सकते हैं। आमतौर पर उपयोग किए जाने वाले चार्ट में बार चार्ट, कॉलम चार्ट, पाई चार्ट और लाइन चार्ट शामिल हैं। ये सभी चार्ट कुछ विविधता (वेरिएंट्स/स्टाइल में बदलाव) में उपलब्ध हैं और एक्सेल के साथ प्लॉट किए गए मेनू के विकल्पों को चुनकर इसका उपयोग किया जा सकता है ।

बार चार्ट और कॉलम चार्ट का उपयोग अक्सर सारांश मानों को चित्रित करने के लिए किया जाता है, जैसे श्रेणीबद्ध चर के मामले में कुल, औसत या आवृत्ति। यहाँ एक उदाहरण है।

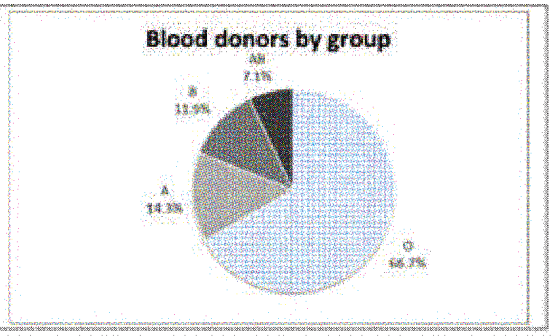
आयुवर्ग	मोटापे से ग्रस्त व्यक्ति
< 30 yrs	21
31-40	42
>= 41 yrs	15



चित्र 9.2: बार चार्ट

एक **पाई चार्ट** का उपयोग किसी विशेषता के विभिन्न घटकों (जैसे प्रतिशत) को प्रदर्षित करने के लिए किया जाता है। रक्त समूह के अनुसार रक्त दाताओं का मासिक प्रतिशत बार चार्ट की तुलना में पाई चार्ट द्वारा बेहतर प्रदर्शित होता है। आप ध्यान दें कि प्रतिशत नीचे दिखाए गए अनुसार 100 जोड़ सकते हैं।

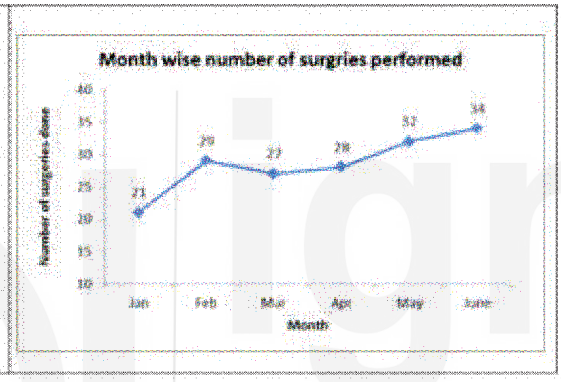
रक्त समूह	रक्त दाताओं
O	56
A	12
B	10
AB	6



चित्र 9.3: पाई चार्ट

एक रेखा चार्ट का उपयोग एक अध्ययन चर के रुझान (समय के साथ बदलते पैटर्न) को समझाने के लिए किया जाता है। उदाहरण के लिए, एक अस्पताल में की गई सर्जरी की संख्या में वृद्धि को नीचे दिखाए गए अनुसार एक लाइन चार्ट द्वारा प्रदर्शित किया जाता है।

Month	# of surgeries
Jan	21
Feb	29
Mar	27
Apr	28
May	32
June	34



चित्र 9.4: लाइन चार्ट

एक्सेल, नवजात शिशुओं के जन्म के समय वजन और पैर की लंबाई जैसे दो चर के बीच संबंधों को समझने में भी मदद करता है। इसे स्कैट आरेख कहा जाता है जोकि इकाई 10 के चित्र 10.3 और 10.1 में दिखाया गया है।

9.4 एक्सेल के साथ सरल सांख्यिकीय विश्लेषण

एक्सेल में डेटा विश्लेषण करने के लिए एक अंतर्निहित सांख्यिकीय पैकेज है। इसे डेटा मेनू में दिया गए विश्लेषण (एनालिसिस) टूलपैक के नाम से जाना जाता है। यह उपकरण मेनू का एक भाग है लेकिन आमतौर पर मेनू में प्रदर्शित नहीं होता है। हालाँकि, इसे एक्सेल विकल्पों में उपलब्ध ऐड-इन्स के माध्यम से सक्रिय किया जा सकता है। यहाँ बुनियादी सांख्यिकीय विश्लेषण के लिए इस उपकरण में उपलब्ध सुविधाओं की एक आंशिक सूची है।

• एनोवा: एकल कारक	• मूविंग एवरेज
• एनोवा: प्रतिकृति के साथ दो-कारक	• प्रतिगमन
• सह-संबंध	• नमूना लेना
• वर्णनात्मक आँकड़े	• टी-परीक्षण
• हिस्टोग्राम	• जेड-परीक्षण
• रैंडम नंबर जनरेशन	

हालांकि, इनमें से प्रत्येक उपकरण को इसके अनुप्रयोग के कुछ सांख्यिकीय ज्ञान की आवश्यकता होती है। जिसकी सहायता से कई गणितीय, वित्तीय, तार्किक और अन्य कार्यों का उपयोग करना भी संभव है।

सरल सांख्यिकीय परीक्षण करने में एक्सेल की उपयोगिताओं को सीखने के लिए यहां दो चित्र दिए गए हैं।

Group-A	20.43	22.51	18.99	20.49	23.12	25.63	18.08	20.63	22.55	22.43	22.77	23.23
Group-B	17.7	21.4	20.7	19.3	21	17.9	18.6	18.5	18.2	20.3		

हम परीक्षण करना चाहते हैं कि क्या दोनों समूहों के बीच औसत बीएमआई का अंतर महत्वपूर्ण है।

समाधानः यदि हम इकाई 8 में दिए गए सूत्र का उपयोग करते हैं, तो हम निम्नानुसार आगे बढ़ते हैं:

समूह ए के लिए हमारे पास n1 = 12, $\overline{X}$ 21-734 और s1 = 2.078 है। समूह बी n2 =10, $\overline{X}$ = 19.36 और s2 = 1-378 के लिए संयुक्त एसडी s = 1.78 होगा।

शून्य परिकल्पना H0: अंतर = 0 और H1: अंतर शून्य नहीं है (दो तरफा परिकल्पना)। अब नमूना डेटा से हम प्राप्त करते हैं

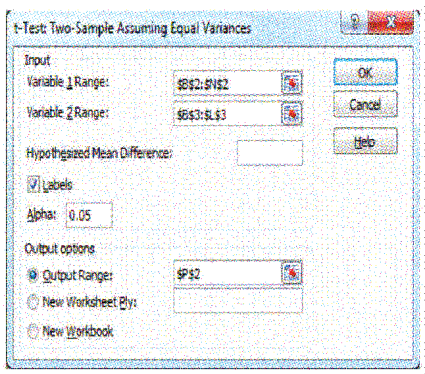
ए) अंतर = 21.734 -19.36 = 2.378

बी) स्टैंडर्ड एरर = $1.797\sqrt{\left\{\frac{1}{12}+\frac{1}{10}\right\}}$ =1.797 * 0.4282 = 0.7694

सी) टेस्ट वैल्यू (t) = 2.378 / 0.7694 = 3.0907

स्वतंत्रत चर (डिग्री ऑफ फ्रीडम) (12. 10-2) = 20 हैं। स्वतंत्रता की 20 डिग्री के लिए तालिकाओं से क्रांतिक मान α = 0.05 लेना 2.08 है। चूंकि t> 2.08 हम शून्य परिकल्पना को स्वीकार नहीं कर सकते हैं और इसलिए यह अंतर 5% के स्तर पर महत्वपूर्ण है।

चित्र 9.5: टी–टेस्ट विकल्प



तालिका 9.2: टी–टेस्ट आउटपुट

टी–टेस्ट : दो–नमूनों को समान भिन्नता को मानते हुए

	Group-A	*Group-B*
Mean	21.738	19.360
Variance	4.319	1.898
Observations	12	10
Pooled Variance	3.229	
Hypothesized Mean Difference	0	
Df	20	
t Stat	**3.090**	
P(T<=t) one-tail	0.002	
t Critical one-tail	1.724	
P(T<=t) two-tail	**0.005**	
t Critical two-tail	2.085	

एक्सेल के साथ काम करनाः वैकल्पिक रूप से, हम डेटा मेनू में एक्सेल के विश्लेषण टूलपैक का उपयोग कर सकते हैं। हमें विकल्प टी-टेस्ट मिलता हैः दो-नमूनों को समान भिन्नता को मानते हुए । दो समूहों के डेटा को उचित शीर्षकों के साथ दो अलग कॉलमों (या पंक्तियों) के रूप में दर्ज किया जा सकता है। ऑप्शन विंडो और आउटपुट को फिगर 9.5 में दिखाया गया है और एक्सेल आउटपुट तालिका 9.2 में दिखाया गया है)।

आउटपुट 3.090 (t स्टेट) के रूप में परीक्षण मान दिखाता है और इसकी तुलना दो टेल परीक्षण के लिए 2.085 के क्रांतिक मान के साथ की जाती है। चूंकि परीक्षण मान, क्रांतिक मान से अधिक है, हम शून्य परिकल्पना को अस्वीकार करते हैं और निष्कर्ष निकालते हैं कि साधनों में अंतर महत्वपूर्ण है।

इसके बजाय, हम परीक्षण के पी-मूल्य का उपयोग कर सकते हैं, चित्र 9.5 और तालिका 9.2 में पी (टी <= टी) दो-टेल के रूप में दिखाया गया है जो 0.002 है। चूंकि यह मान 0.05 (महत्व का 5% स्तर) से काफी कम है, इसलिए हम शून्य परिकल्पना को अस्वीकार करते हैं और अंतर को महत्वपूर्ण मानते हैं।

यहां एक अलग प्रकार के टी-टेस्ट पर एक और दृष्टांत दिया गया है जिसे युग्मित (पैयर्ड) टी-टेस्ट कहा जाता है।

उदाहरण 9.3 (पैयर्ड टी-टेस्ट): निम्नलिखित डेटा 15 व्यक्तियों के उपचार से पहले और 30 दिनों के बाद उपवास रक्त शर्करा (FBS) को संदर्भित करता है। परीक्षण करें कि क्या एफबीएस के स्तर में काफी कमी आई है।

Patient No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Before	176	142	148	130	196	177	162	194	152	156	140	173	192	148	180
After	154	122	127	110	171	157	141	164	136	133	128	148	157	130	166

हम परीक्षण करना चाहते हैं कि क्या एफ.बी.एस. में परिवर्तन सांख्यिकीय रूप से महत्वपूर्ण है।

समाधानः हम एक्सेल के विश्लेषण टूलपैक का उपयोग करें और टी-टेस्ट का चयन करें: मीन के लिए पेयर टू-सैंपल। आउटपुट निम्न परिणाम दिखाता है।

t-test: Paired Two Sample for Means		
	Before	*After*
Mean	164.4	142.9333
Variance	449.6857	330.3524
Observations	15	15
Pearson Correlation	0.9676	
Hypothesized Mean Difference	0	
Df	14	
t Stat	14.2325	
P(T<=t) one-tail	5.09E-10	
t Critical one-tail	1.7613	
P(T<=t) two-tail	1.02E-09	
t Critical two-tail	2.1448	

आप पाएंगे कि टू-टेल वाले परीक्षण का पी-मूल्य वैज्ञानिक संकेतन में as1.02E-09 दिया

गया है। यह 0.00000000102 के बराबर है जो 0.05 से बहुत कम है और इसलिए FBS में परिवर्तन सांख्यिकीय रूप से महत्वपूर्ण है। जब पी-मान इस तरह से छोटा होता है, तो हम बस सभी अंकों को प्रदर्शित करने के बजाय P <0-001 के रूप में रिपोर्ट करते हैं।

एक्सेल की एक अन्य महत्वपूर्ण उपयोगिता दो से अधिक स्वतंत्र समूहों के साधनों की तुलना के लिए एनालिसिस ऑफ वेरिएंस (एनोवा) नामक उपकरण है।

यहाँ एक उदाहरण है।

उदाहरण 9.4 (ANOVA): मान लीजिए कि हम तीन स्वतंत्र टीमों ए, बी और सी द्वारा मापे गए एनीमिक व्यक्तियों के बी 12 स्तर (प्रति मिलीलीटर या मिलीग्राम/पीजी/एमएल) की तुलना करना चाहते हैं। निम्नलिखित तालिका एक चित्रण डेटा दिखाती है।

A	101	92	97	102	115	98	125	101		
B	110	108	180	125	132	147				
C	313	181	252	173	345	197	241	223	250	257

हम तीन समूहों के बीच माध्य मानों की तुलना करना चाहते हैं और सांख्यिकीय महत्व के 5% स्तर पर परीक्षण करना चाहते हैं। तकनीक को एक-कारक A NOVA या एकल-कारक A NOVA कहा जाता है क्योंकि 24 मानों वाले कुल डेटा को तीन स्तरों A, B और C के साथ टीम नामक एकल कारक के अनुसार विभाजित या समूहीकृत किया जाता है।

प्रक्रिया: एक्सेल में, डेटा को तीन अलग-अलग कॉलम या पंक्तियों में उपयुक्त शीर्षकों के साथ दर्ज करना होगा। फिर एनालिसिस टूलपैक खोलें और टूल एनोवा सिंगल फैक्टर को चुनें और ओके दबाएं। विकल्प विंडो डेटा लिंक दिखाती है और आउटपुट नीचे दिखाया गया है।

ANOVA: Single Factor						
SUMMARY						
Groups	Count	Sum	Average	Variance		
A	8	831	103.875	116.125		
B	6	802	133.6667	724.2667		
C	10	2432	243.2	2988.178		
ANOVA						
Source of Variation	SS	Df	MS	F	p-value	F crit
Between Groups	96473.82	2	48236.91	32.33469	3.88E-07	3.4668
Within Groups	31327.81	21	1491.8			
Total	127801.6	23				

उपरोक्त तालिका को एनोवा तालिका कहा जाता है और हमेशा इसी प्रारूप में दिखाई देता है जैसा कि ऊपर दिखाया गया है। साधनों के बीच तुलना समूहों और समूहों के बीच B12 भिन्नता की तुलना द्वारा की जाती है। शून्य परिकल्पना यह है कि सभी समूह साध

ान समान हैं। 'समूह के बीच' दिखाया गया एफ-मूल्य, परीक्षण मूल्य को दर्शाता है। यहां संबंधित पी-मान 0.00000039 है जो 0.05 से कम है। इसलिए B12 मान तीन समूहों के बीच काफी भिन्न होता है।

हम एक्सेल पर चर्चा को इस नोट के साथ समाप्त करते हैं कि 'करके सीखना' एक्सेल में काम करने का और ज्ञान प्राप्त करने का सबसे अच्छा तरीका है। डेटा विश्लेषण के लिए एक्सेल के उपयोग पर अधिक विवरण संदर्भों में दिए गए सरमा (2010) को पढ़ा जा सकता है।

9.5 एस.पी.एस.एस. (SPSS) में डाटा फाइल बनाना

SPSS (सामाजिक विज्ञान के लिए सांख्यिकीय पैकेज) सांख्यिकीय गणना और उन्नत विश्लेषण करने के लिए सॉफ्टवेयर है। मूल रूप से SPSS Inc- द्वारा 50 साल पहले जारी किए गए इस सॉफ्टवेयर में कई सुधार हुए हैं और सर्वेक्षण वैज्ञानिकों, विपणन प्रबंधकों और सामाजिक विज्ञान शोधकर्ताओं के अलावा स्वास्थ्य शोधकर्ताओं के बीच इसकी लोकप्रियता पाई जाती है।

2009 में आईबीएम कॉर्पोरेशन द्वारा इसके अधिग्रहण के बाद, वर्तमान संस्करण को आईबीएम SPSS के रूप में जाना जाता है और वर्तमान संस्करणों को IBM SPSS के रूप में लेबल किया जाता है। इस चर्चा में संस्करण 20 का उपयोग किया जा रहा है।

सामान्य सांख्यिकीय अनुप्रयोगों के लिए एसपीएसएस (SPSS) की कुछ विशेषताएं यहां दी गई हैं।

- हम एसपीएसएस में एक्सेल डेटा खोल सकते हैं।
- विश्लेषण के दौरान एक बार में कई चर संभाले जा सकते हैं। उदाहरण के लिए, सरल फ्रीक्वेंसी टेबल या वर्णनात्मक आँकड़े जैसे माध्य और कई चर के लिए मानक विचलन की गणना एक ही चरण में की जा सकते हैं।
- काई-स्क्वायर परीक्षणों के साथ दो आयामी और बहुआयामी तालिकाओं को निर्मित किया जा सकता है।
- टी-टेस्ट, एनोवा जैसे सांख्यिकीय परीक्षणों को एक बार में केवल चर और समूहों (तुलना किए जाने के लिए) का चयन करके कई चर के लिए किया जा सकता है।
- मल्टीवेरिएट विश्लेषण, पुनरावर्ती युक्तियों के विश्लेषण (repeated measures analysis), चरणबद्ध रैखिक प्रतिगमन जैसी उन्नत सांख्यिकीय विशेषताएं आसानी से नियंत्रित की जा सकती हैं।
- बॉडी मास इंडेक्स जैसे निरंतर चर को श्रेणियों में कोडित किया जा सकता है। मौजूदा श्रेणियों को सरल मेनू विकल्पों के साथ फिर से वर्गीकृत किया जा सकता है।
- हम ऑप्शन विंडो से माउस क्लिक द्वारा सरल सूत्र (कोड नहीं) लिखकर नए चर बना सकते हैं।
- एसपीएसएस डेटा सेटअप में, प्रत्येक पंक्ति को एक केस कहा जाता है। रेकॉर्ड के कुछ हिस्सों (पंक्तियों) का चयन रैंडम तरीके से या किसी ज्ञात नियम (जैसे जेंडर = ''एफ'' और आयु <40 साल) के आधार पर किया जा सकता है।
- जटिल परिस्थितियों (जैसे समूह वार हिस्टोग्राम) के साथ चार्ट दो-आयामी और त्रि-आयामी दृश्य के लिए उपलब्ध हैं।

इन सबसे ऊपर किसी प्रोग्राम को चलाने के लिए कोई कोड लिखने की आवश्यकता नहीं है। वास्तव में, सभी ऑपरेशन माउस क्लिक पर आधारित होते हैं लेकिन फिर भी बैक-एंड कोड को सिंटैक्स के रूप में सहेजा जा सकता है।

SPSS की एक महत्वपूर्ण उपयोगिता आउटपुट है। अधिकांश अनुप्रयोगों के लिए आउटपुट स्वरूपित तालिकाओं के रूप में प्रकट होता है ताकि हम आउटपुट की सामग्री को फिर से बनाए बिना एमएस-वर्ड या एमएस-एक्सेल में समान कॉपी और पेस्ट कर सकें।

जब आउटपुट में कई ऑब्जेक्ट्स (जैसे टेबल या चार्ट) होते हैं, तो पूरा आउटपुट फाइल मेनू में नेपदह एक्सपोर्ट 'विकल्प का उपयोग करके पीडीएफ या वर्ड जैसे अन्य प्रारूप में निर्यात किया जा सकता है।

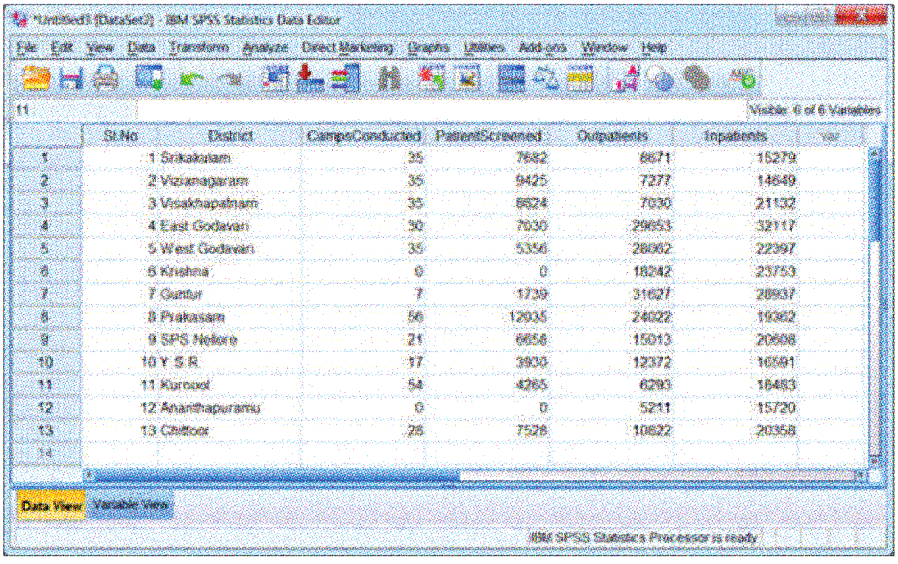
हम हेल्थ कैंप डेटा (तालिका 9.1) को याद करें और एसपीएसएस में एक्सेल फाइल खोलें। निम्नलिखित कदम हैं:

1) एसपीएसएस खोलें और निम्नलिखित विकल्पों पर क्लिक करें
2) फाइल ⟶ ओपन ⟶ चूज टाइप एक्सल
3) फाइल को उसके स्थान से चुनें (यहाँ यह डेस्कटॉप है)। सभी फाइल देखने के लिए उपलब्ध होंगे
4) उस शीट पर क्लिक करें जहां डेटा स्थित है
5) ओके दबाएं।

SPSS विंडो में एक्सेल डेटा जैसा कि चित्र 9.6 में दिखाया गया है।

आइए हम डेस्कटॉप पर हेल्थ कैंप डेटा के रूप में फाइल को सहेजते हैं। फाइल को एक्सटेंशन 'ssav' के साथ सहेजा जाएगा जबकि Excel में उसी फाइल में एक्सटेंशन '.xls' था।

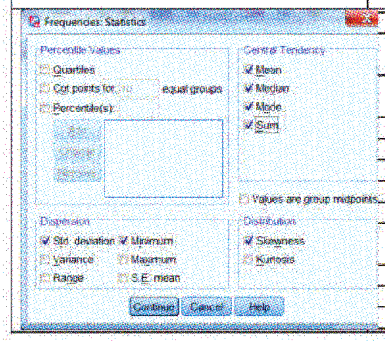
प्रत्येक डेटा में दो व्यूज होते हैं। डेटा व्यू और वेरिएबल व्यू, जैसा कि चित्र 9.6 में दिखाया गया है। 'परिवर्तनीय दृश्य' में हम चर और उनके गुणों के नाम पाते हैं। यह ध्यान रखना महत्वपूर्ण है कि SPSS यह निर्दिष्ट करने के लिए कहता है कि एक चर संख्यात्मक या व्याख्यात्मक है। हमें डेटा का प्रकार (नाममात्र, क्रमिक या पैमाना) भी निर्दिष्ट करना चाहिए। कुछ सांख्यिकीय परीक्षण डेटा के प्रकार पर निर्भर करते हैं।



चित्र 9.6: एसपीएसएस में खोला गया एक्सेल डेटा

आइए हम हेल्थ कैंप डेटा के लिए सारांश आँकड़े देखें। ऐसा करने के लिए, विकल्प का चयन करें Analyze → Descriptive Statistics → Frequencies। परिणामी मेनू में हम उन चर का चयन कर सकते हैं जिनके लिए वर्णनात्मक आंकड़ों की आवश्यकता होती है। हम वैकल्पिक रूप से आउटपुट के रूप में औसत, मानक विचलन, न्यूनतम, अधिकतम आदि जैसे सारांश मानों का चयन कर सकते हैं।

चित्र 9.7: सांख्यिकी विकल्प



तालिका 9.3: एसपीएसएस आउटपुट

Statistics

		Patient Screened	Out-patients	In-patients
N	Valid	13	13	13
	Missing	0	0	0
Mean		5713.23	15715.00	20722.00
Median		6658.00	12372.00	20358.00
Mode		0	5211[a]	14649[a]
Std. Deviation		3634.234	9601.497	5204.917
Skewness		-.204	.607	1.028
Std. Error of Skewness		.616	.616	.616
Minimum		0	5211	14649
Sum		74272	204295	269386

a. Multiple modes exist. The smallest value is shown

चित्र 9.7: विकल्प विंडो दिखाता है। अगर हम तीन वेरिएबल्स (चरों) का चयन करते हैं, तो मरीजों की जांच, आउट-मरीज और इन-मरीजों का आउटपुट तालिका 9.3 में दिखाए गए के समान होगा।

जब कोई तालिका एसपीएसएस से वर्ड (Word) में कॉपी की जाती है, तो उसे वर्ड टेबल्स में उपलब्ध विकल्पों के साथ उपयुक्त रूप से स्वरूपित किया जा सकता है। एसपीएसएस एक व्यक्तिगत मान की संख्या की गणना करके आवृत्ति तालिकाओं को उत्पन्न करता है। हम सीधे वर्ग अंतराल के साथ एक तालिका प्राप्त नहीं करते हैं। डेटा में असामान्य मान की पहचान करने के लिए यह बहुत उपयोगी है.

9.6 क्रॉस टेबुलेशन और काई स्क्वायर टेस्ट

हम SPSS का उपयोग करके अपक्व (Raw) डेटा से दो आयामी तालिकाओं का निमार्ण कर सकते हैं। उदाहरण के लिए, हम पुरुष और महिला रोगियों के बीच मधुमेह (हां/नहीं) के मामलों की संख्या की गणना करना पसंद कर सकते हैं। छोटे डेटा सेटों के लिए हम हाथ से ही गणना कर सकते हैं लेकिन हजारों रिकॉर्ड के साथ हमें सॉफ्टवेयर की आवश्यकता होती है और एसपीएसएस में इस कार्य को करने के लिए क्रॉस्टैब्स नामक मॉड्यूल होता है। गणना की तालिका बनाने के अलावा, हमारे पास स्पष्ट विशेषताओं के बीच परिकल्पना का परीक्षण करने के लिए विकल्प हैं।

यहाँ एक उदाहरण है।

उदाहरण 9.5: स्वास्थ्य देखभाल अध्ययन में 30 रोगियों के विवरण वाले एक्सेल फाइल में बनाए गए चित्रण डेटा का एक हिस्सा निम्नलिखित है।

तालिका 9.4: 30 मामलों और 12 चर के साथ चित्रण डेटा

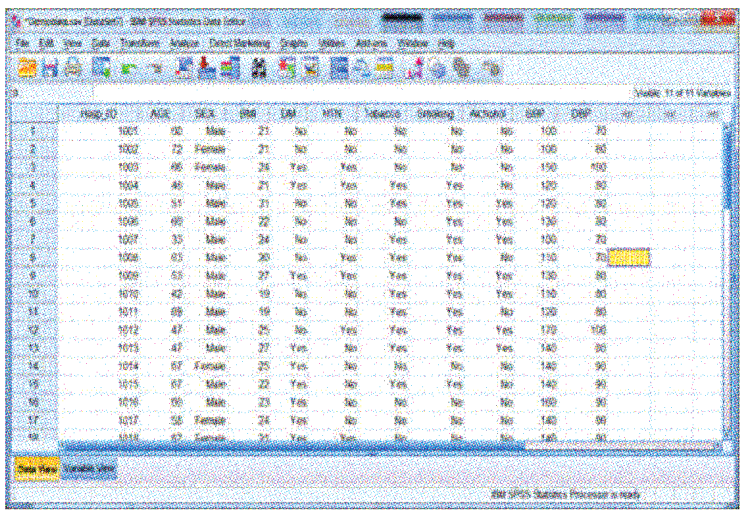
Hosp_ID	AGE	SEX	BMI	DM	HTN	Tobacco	Smoking	Alcohol	SBP	DBP
1001	60	1	20.76	0	0	0	0	0	100	70
1002	72	2	21.4	0	0	0	0	0	100	60
1003	66	2	23.5	1	1	0	0	0	150	100
1004	46	1	21.1	1	1	1	1	0	120	80
1005	51	1	30.61	0	0	1	1	1	120	80
1006	60	1	21.78	0	0	0	1	1	130	80

1007	33	1	24.11	0	0	1	1	1	100	70
1008	63	1	20.2	0	1	1	1	0	110	70
1009	53	1	26.77	1	1	1	1	1	130	80
1010	42	1	19.15	0	0	1	1	1	110	80
1011	69	1	19.49	0	`0	1	1	0	120	80
1012	47	1	25.28	0	1	1	1	1	170	100
1013	47	1	26.56	1	0	1	1	1	140	80
1014	67	2	24.8	1	0	0	0	0	140	90
1015	67	1	21.8	1	0	1	1	0	140	90
1016	60	1	22.86	1	0	0	0	0	160	90
1017	58	2	23.56	1	0	0	0	0	140	90
1018	67	2	20.81	1	1	0	0	0	140	90
1019	59	1	24.03	1	1	0	0	0	120	80
1020	45	2	25.68	0	1	0	0	0	120	80
1021	49	1	24.22	0	1	0	0	0	150	90
1022	65	1	18.73	1	1	1	1	1	160	90
1023	40	2	28.54	1	1	0	0	0	120	80
1024	41	1	26.22	0	1	0	0	0	150	70
1025	55	2	19.63	1	0	0	0	0	130	80
1026	68	2	25.97	1	1	0	0	0	140	90
1027	45	1	22.86	1	0	1	1	1	110	80
1028	60	1	20.76	0	0	0	0	0	110	70
1029	60	1	28.89	0	1	1	1	1	160	90
1030	60	1	18.31	0	1	1	1	0	140	80

निम्नलिखित कोड का उपयोग किया जाता है।

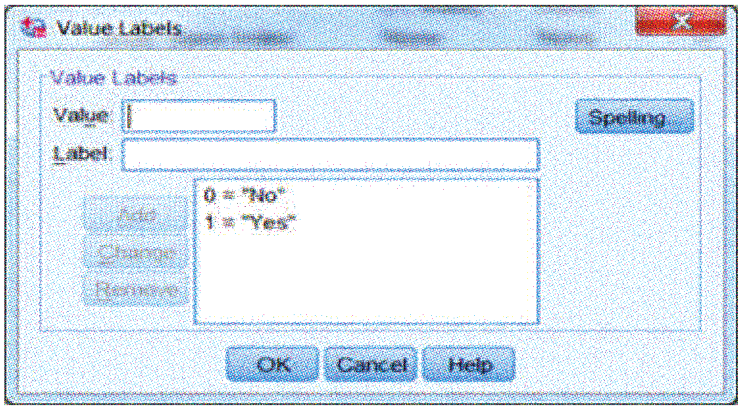
1) सेक्स (1 = पुरुष, 2 = महिला)
2) डीएम = मधुमेह (1 = हां, 0 = नहीं)
3) HTN = उच्च रक्तचाप (1 = हाँ, 0 = नहीं)
4) तम्बाकू = तम्बाकू चबाना (1 = हाँ, 0 = नहीं)
5) धूम्रपान (1 = हाँ, 0 = नहीं)
6) शराब (1 = हाँ, 0 = नहीं)

यह फाइल SPSS में खोली जा सकती है और कोड ऊपर बताए अनुसार दिए गए हैं। डेटा फाइल आंकड़ा 9.8 में दिखाया गया है।



चित्र 9.8: कोड के लिए लेबल के साथ डेटा फाइल

लिंग, मधुमेह या उच्च रक्तचाप जैसे चरों के लिए हमें संख्यात्मक कोड के लिए लेबल असाइन करना होगा। यह डेटा फाइल के 'वैरिएबल व्यू' में किया जा सकता है। 'मान' शीर्षक वाला एक स्तंभ है और हमें इस कॉलम में चर (डीएम) के अनुसार क्लिक करना होगा। लेबल को चित्र 9.9 में दिखाया गया है।



चित्र 9.9: न्यूमेरिक कोड के लिए लेबल असाइन करने के विकल्प

फाइल के डेटा दृश्य में, अगर हम View ⟶ Value लेबल पर क्लिक करते हैं, तो वास्तविक लेबल के बजाय संख्यात्मक कोड दिखाई देते हैं। अब हम निम्नलिखित प्रश्नों का उत्तर देते हैं। मामलों के कोडिंग और चयन पर कुछ संबंधित विवरण संदर्भों में दिए गए (सरमा, 2010) हो सकते हैं।

अपनी प्रगति जांचें

4) एक्सेल के विश्लेषण टूलपैक में कौन कौन से सांख्यिकीय उपकरण उपलब्ध हैं? टी-टेस्ट विकल्पों के बारे में लिखें।

.....

.....

.....

5) SPSS में क्रॉस टेबुलेशन का क्या मतलब है? एक उदाहरण दें।

.....

.....

.....

प्रश्न 9.1: उन रोगियों की गिनती करें जो धूम्रपान करते हैं और जिन्हें मधुमेह भी है।

Analyze ⟶ Descriptive Statistics ⟶ Crosstabs को चुनकर किया जाता है। यह एनालाइज परिणामी विंडो में डीएम (मधुमेह) को पंक्तियों (Row) में भेजें और स्तंभों (Column)के लिए धुम्रपान (SMOKING) का चुनाव करें। ओके दबाएं। यह काउंट्स की निम्न दो-तरफा तालिका (प्रत्येक श्रेणी में रोगियों की संख्या) देता है। आपको प्रतिशत रिपोर्ट करने के लिए विकल्प मिलेंगे (सेल बटन पर क्लिक करें)।

Diabetes * Smoking Cross tabulation			
Count			
Diabetes	Smoking		Total
	No	Yes	
No	6	9	15
Yes	9	6	15
Total	15	15	30

प्रश्न 9.2: क्या तम्बाकू चबाने और शराब पीने के बीच कोई संबंध है? तंबाकू और शराब के बीच संबंध (एसोसिएशन) के संदर्भ में व्यक्त किए जाएंगे क्योंकि दो चर गुणात्मक (श्रेणीबद्ध) हैं। हम शून्य परिकल्पना का परीक्षण कर सकते हैं कि तंबाकू और शराब एक दूसरे संबंधित नहीं हैं। यदि इस परिकल्पना को खारिज कर दिया जाता है, तो यह मानने का कारण होगा कि वे एक दूसरे संबंधित नहीं हैं, जिसका अर्थ है कि एक दूसरे संबंधित होने की संभावना है। अब काई स्क्वायर टेस्ट क्रॉसस्टैब्स में एक विकल्प है। यदि आप क्रॉसस्टैब्स विंडो में 'सांख्यिकी बटन पर क्लिक करते हैं, तो आप 'काई स्क्वायर' के लिए जाँच करेंगे। यह निम्न तालिका फिशर का सटीक परीक्षण पी-मान = 0.001 (2-पक्षीय सटीक सार्थकता के रूप में चिह्नित है) दिखाता है। चूंकि पी-मान 0.05 से छोटा है, इसलिए हम स्वतंत्रता की शून्य परिकल्पना को अस्वीकार कर सकते हैं और निष्कर्ष निकाल सकते हैं कि तंबाकू और शराब की खपत प्रत्येक के साथ जुड़ी हुई है।

Tobacco Chewing	Alcohol		Total
	No	Yes	
No	15	1	16
Yes	5	9	14
Total	20	10	30

प्रश्न 9.3: पुरुष और महिला रोगियों के बीच एसबीपी और डीबीपी का औसत और मानक विचलन क्या है? चूंकि एसबीपी (Systolic Blood Pressure) और डीबीपी (Diastolic Blood Pressure) निरंतर हैं (स्पष्ट नहीं) हम औसत और मानक विचलन पा सकते हैं। विकल्प हैं। Analyze → Compare Means → Means विकल्प विंडो के भीतर SBP और DBP को 'निर्भर सूची' में चुनेंय 'स्वतंत्र सूची' बॉक्स में सेक्स। विकल्पों को छोड़ दें जैसे वे हैं और ओके दबाएं। यह निम्नलिखित परिणाम देता है।

SEX		Systolic Blood Pressure	Diastolic Blood Pressure
Male	Mean	130.95	80.95
	N	21	21
	Std. Deviation	21.425	8.309
Female	Mean	131.11	84.44
	N	9	9
	Std. Deviation	15.366	11.304
Total	Mean	131.00	82.00
	N	30	30
	Std. Deviation	19.538	9.248

औसत और मानक विचलन को लिंग के आधार पर और समग्र डेटा के लिए दिया जाता है। (कुल 'समग्र' नमूने (लिंग/सेक्स के बिना) को इंगित करता है। आपको यह मान प्राप्त करने के लिए साधन नहीं जोड़ना चाहिए।

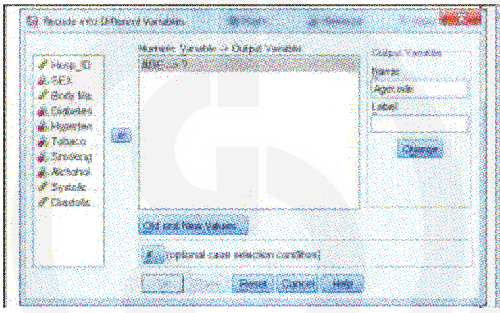
यहाँ एक और उदाहरण है।

उदाहरण 9.6: उदाहरण 9.3 में उपयोग किए गए डेटा पर पुनर्विचार करें। आइए हम आयु को तीन समूहों में विभाजित करें: i) 50 वर्ष से कम, ii) 51-60 वर्ष और iii) 61 और उससे अधिक। तब हम एसबीपी या डीबीपी आयु वार के माध्य की तुलना कर सकते हैं।

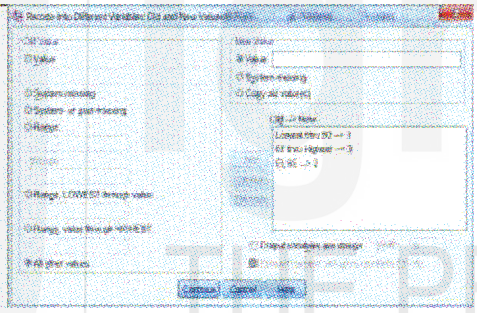
दृष्टिकोण: उम्र को वर्गीकृत करने के लिए, हम मेनू विकल्प ट्रांसफॉर्म → रिकोड को विभिन्न चर (Recode into Different Variable) में उपयोग करते हैं। विकल्प विंडो में आयु चुनें और बीच में दिखाए गए बॉक्स में भेजें।

फिर 'चर कोड' के रूप में नए चर के लिए एक नाम असाइन करें। फिर श्परिवर्तनश पर क्लिक करें और फिर 'पुराने और नए मूल्यों' (वैल्यू) पर क्लिक करें। प्रत्येक श्रेणी के लिए, 1, 2, 3 जैसी संख्या निर्दिष्ट करें और हर बार 'Add' दबाएँ। ये विकल्प चित्र 9.8 में दिखाए गए हैं।

एक बार हो जाने के बाद, आप OK बटन पर क्लिक कर सकते हैं। यह एक नया वैरिएबल 'एज कोड' बनाता है और संबंधित कोड वाला एक नया कॉलम जेनरेट होता है। फिर (1= labels '50 वर्ष से कम', 2 = 51-60 वर्ष', 3 = 61 और उससे अधिक) के कोड को लेबल असाइन करें।



चित्र 9.10 (क): रीकोडिंग के लिए चर का चयन



चित्र 9.10 (बी): पुराने मानों को नए कोड में बदलना

यह रीकोडिंग प्रक्रिया को पूरा करता है। हम प्रश्न 3 के उत्तर में दिए गए विकल्पों का उपयोग करके SBP और DBP के माध्य मान ज्ञात कर सकते हैं। समूह वार माध्य तालिका 9.5 में दिखाए गए हैं।

Table 9.5: Descriptive Statistics of SBP and DBP by Age Group

Age	Systolic Blood Pressure			Diastolic Blood Pressure		
	Mean	N	Std. Deviation	Mean	N	Std. Deviation
<= 50 Years	129.00	10	22.336	81.00	10	8.756
51-60 Years	130.91	11	18.684	80.91	11	7.006
61 & Above	133.33	9	19.365	84.44	9	12.360
Total	131.00	30	19.538	82.00	30	9.248

चूंकि हमने आयु समूहों के लिए लेबल संलग्न किए हैं, इसलिए हम शीर्षक को आयु के रूप में लिखते हैं न कि आयु कोड के रूप में। 'pivot options' का उपयोग करके 'एसपीएसएस द्वारा निर्मित मूल तालिका आउटपुट में स्वयं बदल जाती है, (SPSS आउटपुट में तालिका पर डबल क्लिक करके देखें)।

हम एक्सेल और एसपीएसएस की मूल बातों की चर्चा को इस अवलोकन के साथ समाप्त करते हैं कि एक्सेल सांख्यिकीय डेटा को संभालने के लिए प्रवेश स्तर के सॉफ्टवेयर की तरह कुछ कार्यों को करता है जबकि एसपीएसएस व्यापक तरीके से इसमें मदद करता है।

9.7 सारांश

- हमने देखा है कि विभिन्न प्रकार के अनुसंधान डेटा को संभालने के लिए एक्सेल शोधकर्ताओं के बीच लोकप्रिय उपकरण है। एक्सेल वर्कशीट का उपयोग करके मध्यम से बड़े आकार के डेटा सेट बनाए जा सकते हैं। डेटा को प्रभावी ढंग से प्रबंधित करने और गणना करने के लिए कई संपादन और कंप्यूटिंग सुविधाएं हैं। बार चार्ट, पाई चार्ट और लाइन चार्ट जैसे सांख्यिकीय चित्र प्रभावी रूप से एक्सेल का उपयोग करके बनाए जा सकते हैं।
- हम औसत, मानक विचलन, सहसंबंध खोजने और परिकल्पना के सरल परीक्षण करने जैसे कुछ सांख्यिकीय विश्लेषण भी कर सकते हैं।
- हमने यह भी सीखा है कि एसपीएसएस विशेष रूप से बुनियादी और उन्नत सांख्यिकीय विश्लेषणों को संभालने के लिए है। हम एसपीएसएस में एक्सेल डेटा फाइलें खोल सकते हैं और डेटा में बदलाव (जैसे रीकोडिंग, लेबल संलग्न करना) के अलावा विश्लेषण कर सकते हैं। संक्षेप में, एक्सेल और एसपीएसएस दोनों सार्वजनिक स्वास्थ्य शोधकर्ताओं के लिए उपयोगकर्ता के अनुकूल हैं।

9.8 संदर्भ

सरमा, के.वी.एस. (2010). स्टटिस्टिक्स मेड सिंपल डू इट योरसेल्फ ऑन पीसी, दूसरा संस्करण. नई दिल्लीः प्रेंटिस हॉल ऑफ इंडिया.

सामाजिक आर्थिक सर्वेक्षण (2016-17) आंध्र प्रदेश सरकार.

9.9 आपनी प्रगति की जाँच करने के लिए उत्तर

1) एक्सेल में डाटा फाइल को वर्कबुक कहा जाता है। इसमें आम तौर पर तीन शीट होती हैं। हर शीट में 16384 कॉलम होते हैं जिन्हें ए, बी, .. और 1048576 पंक्तियों के साथ 1,2,3 गिना जाता है। विवरण के लिए भाग 9.1 देखें।

2) एक्सेल शीट की चार महत्वपूर्ण संपादन विशेषताएं हैंः ए) डेटा चयन, बी) सॉर्ट/फिल्टर, सी) कट, कॉपी और पेस्ट डी) वर्ड को डेटा निर्यात करना। विवरण के लिए भाग 9.2 देखें।

3) एक्सेल में, पहली पंक्ति (हेडिंग रो) रखने के लिए और पहला कॉलम हमेशा दिखाई देने वाला फ्रीज पैन विकल्प मुख्य मेनू से उपयोग किया जाता है। विवरण के लिए भाग 9.2 देखें।

4) एक्सेल के विश्लेषण टूलपैक में कई सांख्यिकीय उपकरण उपलब्ध हैं। उदाहरण एनोवा, सहसंबंध, वर्णनात्मक सांख्यिकी, हिस्टोग्राम, रैंडम नंबर जनरेशन, सैम्पलिंग, टी-टेस्ट, जेड-टेस्ट आदि। अधिक विवरण के लिए भाग 9.4 देखें।

5) हम एसपीएसएस का उपयोग करके अपक्व डेटा से दो आयामी तालिकाओं को उत्पन्न कर सकते हैं। उदाहरण के लिए, हम पुरुष और महिला रोगियों के बीच डीएम (हां/नहीं) के मामलों की संख्या की गणना करना पसंद कर सकते हैं। छोटे डेटा सेटों के लिए हम हाथ के साथ गणना कर सकते हैं लेकिन हजारों रिकॉर्ड के साथ हमें सॉफ्टवेयर की आवश्यकता होती है और एसपीएसएस में इस कार्य को करने के लिए क्रॉस्टैब्स नामक मॉड्यूल होता है। विवरण के लिए भाग 9.6 देखें।

इकाई 10 उन्नत सांख्यिकी*

इकाई की रूपरेखा

अधिगम के उद्देश्य

इस इकाई को पढ़ने के बाद, आप निम्न कार्य कर सकेंगे:

- साहचर्य के विभिन्न मानों की गणना और व्याख्या करना;
- काई स्क्वायर परीक्षण करना;
- रैखिक प्रतिगमन के साथ काम करना; तथा
- प्रसरण विश्लेषण का निर्वचन करना और निष्कर्षों की व्याख्या करना।

10.0 परिचय

अक्सर चिकित्सा और सार्वजनिक स्वास्थ्य अनुसंधान में, हमें *श्रेणीबद्ध* चर के साथ काम करना होता है जिन्हें *गुणात्मक कारकों* के रूप में भी जाना जाता है। ऐसे कारकों पर डेटा एक माप के रूप में नहीं होता है, बल्कि यह संभावित मूल्यों की असतत सूची से चुनने के लिए एक विकल्प की तरह होता है। उदाहरण के लिए, एक गाँव में स्वच्छता स्तर को *निम्न, मध्यम और उच्च* के रूप में व्यक्त किया जा सकता है जिसे क्रमशः 1, 2, 3 के रूप में कोडित किया जा सकता है। इस प्रकार के डेटा को क्रमिक डेटा कहा जाता है क्योंकि इसमें क्रम (ऑर्डर) या विकल्पों का एक अर्थ होता है। एक उच्च मूल्य एक बेहतर स्थिति को इंगित करता है। कभी-कभी श्रेणियां सिर्फ इस मायने में नाममात्र की होती हैं कि संख्यात्मक मान किसी क्रम (ऑर्डर) को इंगित नहीं करता हैं।

उदाहरण के लिए, 4 श्रेणियों के तहत कुष्ठरोग के प्रकार को 1, 2, 3, 4 के रूप में कोडित किया जा सकता है और कोड 1 के साथ तुलना करने पर कोड 4 बेहतर स्थिति का संकेत माना जाता है।

श्रेणीबद्ध डेटा के लिए हम माध्य और मानक विचलन जैसे उपायों का उपयोग नहीं कर सकते हैं। इसके बजाय इसे मामलों को गिनती और प्रतिशत (या अनुपात) के रूप में व्यक्त करना होता है।

* योगदानकर्ता – प्रो. के.वी.एस. सरमा (सेवानिवृत्त), सांख्यिकी विभाग, श्री वेंकटेश्वर विश्वविद्यालय, तिरुपति.
अनुवादक – डॉ. निशीथ राय, सहायक प्रोफेसर, मानव विज्ञान, म.गां.अ.हिं.वि. वर्धा, महाराष्ट्र ।

जब दो श्रेणीबद्ध चर की तुलना की जानी है, तो हम आंकड़ों को दो-तरफा तालिका के रूप में सारांशित करते हैं, जिसे आसंग तालिका (contingency table) या क्रॉस सारणीकरण के रूप में जाना जाता है।

उदाहरण के लिए, एक अध्ययन क्षेत्र में मोतियाबिंद की व्यापकता का लिंग वार वितरण तालिका 10.1 में दिखाए गए 2 ग 2 तालिका के रूप में प्रस्तुत किया जा सकता है।

तालिका 10.1: लिंग बनाम मोतियाबिंद का क्रॉस सारणीकरण

Gender	Cataract	Total	
	Yes	No	
Male	73	54	127
Female	35	38	73
Total	108	92	200

इस संदर्भ में दो कारक हैं; एक लिंग (पुरुष या महिला) है और दूसरा मोतियाबिंद (हाँ या नहीं) की उपस्थिति है। हम यह जानना चाहते हैं कि क्या मोतियाबिंद का प्रचलन लिंग के साथ कोई संबंध है या मोतियाबिंद लिंग से स्वतंत्र है।

इस प्रकार की आसंग तालिकाओं को एक्सेल और एसपीएसएस के उपकरणों का उपयोग करके आसानी से बड़े डेटा पर तैयार किया जा सकता है। Excel में हम सम्मिलित मेनू से विकल्प 'धुरी तालिका' 'Pivot table' from the insert menu का उपयोग करते हैं।

10.1 काई-स्क्वायर साहचर्य का परीक्षण

साहचर्य (एसोसिएशन) शब्द का उपयोग दो गुणात्मक कारकों के बीच संबंध को इंगित करने के लिए किया जाता है, जिसे विशेषताओं के रूप में भी जाना जाता है। तालिका 10.1 में दिए गए 4 मानों को सामान्य रूप से नीचे दिखाया जा सकता है ताकि हम आगे के विश्लेषण के लिए एक सूत्र तैयार कर सकें।

Gender	Cataract	Total	
	Yes	No	
Male	a	b	(a+b)
Female	c	d	(c+d)
Total	(a+c)	(b+d)	N

यदि हम यह जानना चाहते हैं कि क्या लिंग के साथ मोतियाबिंद की उपस्थिति का कोई साहचर्य (संबंध) है तो हम सार्थकता का एक सांख्यिकीय परीक्षण 'स्वतंत्र काई-स्क्वायर'का उपयोग कर सकते हैं।

शून्य परिकल्पना H0 है: दो विशेषताएँ स्वतंत्र हैं और हम α (0.05 लेते हैं। हम ग्रीक अक्षर द्वारा निरूपित एक परीक्षण मान की गणना निम्नानुसार करते हैं।

$$\chi^2 = \frac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

यदि परिकलित परीक्षण मान महत्वपूर्ण मान (सांख्यिकीय तालिकाओं से पढ़ने के लिए) से अधिक है तो हम H0 को अस्वीकार करते हैं और निष्कर्ष निकालते हैं कि मोतियाबिंद के प्रसार में लिंग के साथ कुछ संबंध है। क्रांतिक मान हालांकि (# पंक्तियों -1) x (# कॉलम -1) द्वारा दिए गये स्वतंत्र अंश से जुड़ा हुआ है। 2x2 तालिका में हमें (2-1) x

(2-1) = 1. काई स्क्वायर मूल्यों की तालिकाओं से, हमें 1 डिग्री की स्वतंत्रता के साथ 5% के महत्व पर 3.84 के रूप में महत्वपूर्ण मूल्य मिलता है। काई स्क्वायर टेस्ट के बारे में अधिक जानकारी सुंदरा राव और रिचर्ड (2012) से पढ़ी जा सकती है ।

उदाहरण 10.1: तालिका 10.1 में दी गई आषंग तालिका पर विचार करें, जिसे कोष्ठक में अंकित a, b, c, d के साथ नीचे के रूप में पुनः प्रस्तुत किया गया है।

Gender	Cataract	Total	
	Yes	No	
Male	73 (a)	54 (b)	127
Female	35 (c)	38 (d)	73
Total	108	92	200

समाधानः उपरोक्त तालिका से हम निम्नलिखित निरीक्षण करते है।

1) (a+b) =108, (c+d)= 92, (a+c)= 127 and (b+d)= 73 and N = 200

2) काई-स्क्वायर $\frac{200*(2774-1890)^2}{(108)(92)(127)(73)} = 1.697$

3) 1 फ्रीडम ऑफ डिग्री(स्वतंत्र डिग्री) के साथ 5% के स्तर पर सांख्यिकीय तालिकाओं से क्रांतिक मान 3.84 प्राप्त है।

चूंकि गणना किए गए काई-स्क्वायर का मान क्रांतिक मान से बहुत छोटा है, इसलिए हम निष्कर्ष निकालते हैं कि लिंग का मोतियाबिंद की उपस्थिति के साथ कोई साहचर्य (संबंध) नहीं है। इसका मतलब है कि मोतियाबिंद की उपस्थिति लिंग से स्वतंत्र होने की संभावना है।

बड़ी तालिकाओं के लिए काई-स्क्वायर परीक्षणः कभी-कभी हमें 2x2 टेबल से बड़े आकार की आसंग तालिकाएं मिलती हैं। यदि एक कारक में 3 स्तर होते हैं और दूसरे में 4 स्तर होते हैं, तो हमें 3 x 4 तालिका मिलती है। इसमें 3 पंक्तियां और 4 कॉलम होंगे। ऐसी तालिकाओं के लिए काई-स्क्वायर मूल्य की गणना में एक सामान्य सूत्र होता है जिसमें हम प्रत्येक कोशिका के लिए अपेक्षित आवृत्ति (ई) पाते हैं और उनकी तुलना सेल की अवलोकन आवृत्ति (ओ) से करते हैं। एक तालिका में सेल और एक पंक्ति और एक स्तंभ के परिच्छेदन का एक हिस्सा है। एक सेल में प्रेक्षित डेटा होता है।

अपेक्षित आवृत्ति आकस्मिक तालिका से मिली है

ई = (कुल पंक्ति * कुल स्तंभ) / कुल योग।

काई-स्क्वायर मान की गणना सूत्र के साथ की जाती है. $\chi^2 = \sum \frac{(O-E)^2}{E}$

$$\chi^2 = \frac{(O_1-E_1)^2}{E_1} + \frac{(O_2-E_2)^2}{E_2} + \ldots\ldots + \frac{(O_k-E_k)^2}{E_k}$$

शून्य परिकल्पना फिर से H0 हैः दो कारक स्वतंत्र हैं (कोई साहचर्य नहीं है)।

3 x 4 तालिका के लिए स्वतंत्रता की डिग्री (3-1) x (4-1) = 2 x 3 = 6 होगी और स्वतंत्रता की इन डिग्री के लिए क्रांतिक मान को देखा जाएगा। यहाँ 3 पंक्तियों और 3 स्तंभों के साथ एक उदाहरण है।

उदाहरण 10.2: 124 स्वास्थ्य कार्यकर्ताओं के एक बैच को स्क्रीनिंग टेस्ट के लिए नए उपकरण का उपयोग करने पर क्षेत्र प्रशिक्षण दिया गया था। प्रदर्शन स्कोर (अधिकतम = 100) और अनुभव (वर्षों में) के अनुसार उत्तरदाताओं का वितरण नीचे दिया गया है।

Experience (years)	Performance score		
	< 40	41 - 60	Above 60
Up to 5	19	11	8
6 – 10	10	15	18
Above 10	6	17	20

हम परीक्षण करना चाहते हैं कि प्रदर्शन स्कोर का स्वास्थ्य कार्यकर्ता के अनुभव के साथ कोई संबंध है या नहीं।

समाधानः हम इस परीक्षण को करने और निम्नलिखित परिणाम प्राप्त करने के लिए एक ऑनलाइन कैलकुलेटर का उपयोग कर सकते हैं। उदाहरण के लिए, यदि हम https://www.socscistatistics.com/tests/chisquare2/default2.aspx. का उपयोग करते हैं।, हम मध्यवर्ती गणना प्राप्त करते हैं। यह हाथ से गणना पद्धति के बजाय एक सुविधाजनक विधि है।

	< 40	41 - 60	Above 60	Row Total
Up to 5	19 (10.73) [6.38]	11 (13.18) [0.36]	8 (14.10) [2.64]	38
6 – 10	10 (12.14) [0.38]	15 (14.91) [0.00]	18 (15.95) [0.26]	43
Above 10	6 (12.14) [3.10]	17 (14.91) [0.29]	20 (15.95) [1.03]	43
Column Total	35	43	20	124

तालिका की प्रत्येक कोशिका में तीन प्रविष्टियाँ निम्नानुसार हैं:

i) साधारण ब्रैकेट में दिखाए गए सेल की अपेक्षित आवृत्ति (ई)।

ii) सेल के लिए काई-स्क्वायर मान और स्क्वायर ब्रैकेट में दिखाया गया आंकड़ा। कोशिकाओं का काई-स्क्वायर मान जोड़ने पर 14.442 के रूप में परीक्षण मूल्य देता है। परीक्षण का पी-मान च = 0.006 है जो 0.05 से कम है। इसलिए स्कोर और अनुभव के बीच एक महत्वपूर्ण संबंध है।

टिप्पणीः काई-स्क्वायर परीक्षण दो स्पष्ट चर के बीच सहयोग के महत्व को मापता है।

10.2 साहचर्य संबंधित मान

महामारी विज्ञान के कुछ अध्ययनों में, हम सापेक्ष जोखिम (रिलेटिव रिस्क) और बाधाओं (बैरियर) जैसे मानों को साथ पाते हैं, जो दोनों काई स्क्वायर टेस्ट में किए गए 2 x 2 तालिका की गणना पर आधारित होती हैं। एक आबादी के लोगों का अनुपात, उन सभी के बीच बीमारी होना जो किसी स्थिति के संपर्क में हैं, बीमारी का यह जोखिम सम्बंधित जोखिम कहलाता है।

मान लीजिए कि 300 पुरुषों के एक अध्ययन समूह में, हमने 156 धूम्रपान करने वालों को पाया और उनमें से 85 को दिल की बीमारी थी। इस समूह में धूम्रपान के कारण हृदय रोग का खतरा 85/156 = 0.54 या 54% होगा। 146 धूम्रपान न करने वालों में से 28 व्यक्तियों को हृदय रोग था। फिर धूम्रपान न करने वालों के लिए बीमारी का खतरा 28/146 = 0.19 या 19% है। इस जानकारी को नीचे दिखाए अनुसार 2x2 तालिका के रूप में व्यवस्थित किया जा सकता है।

Smoking	Disease		
	Present	Absent	Total
Yes	85	71	156
No	28	118	146
Total	113	189	300

एक बीमारी के सापेक्ष जोखिम (रिलेटिव रिस्क RR) की गणना निम्नानुसार की जाती है।

RR = उजागर समूह में बीमारी का खतरा/अप्रकाशित समूह में बीमारी का खतरा

इस मामले में हमें RR = 0.54 / 0-19 = 2.84 मिलता है। इसका मतलब है कि धूम्रपान न करने वालों की तुलना में धूम्रपान करने वालों को हृदय रोग का खतरा 2.84 गुना अधिक होता है।

RR को जोखिम दर भी कहा जाता है और समय-समय या भौगोलिक स्थानों पर विभिन्न स्वास्थ्य स्थितियों के जोखिम या घटनाओं की तुलना करने के लिए लोकप्रिय रूप से उपयोग किया जाता है। आप पिछले वर्ष की तुलना में सितंबर से दिसंबर के दौरान डेंगू बुखार के आरआर को जानना चाह सकते हैं।

विषम अनुपात या (ऑड-रेसियो OR) केस-कंट्रोल अध्ययन के संदर्भ में उपयोग किए गए साहचर्य का एक और मान है। उन व्यक्तियों के एक समूह को जिन्हें रोग कहा जाता है (जिन्हें केस कहा जाता है) को अध्ययन के लिए चुना जाता है और रोग के बिना तुलनीय व्यक्तियों के एक अन्य समूह (जिन्हें नियंत्रण कहा जाता है) को तुलना के लिए चुना जाता है।

जब डेटा को तालिका 10.1 के रूप में प्रस्तुत किया जाता है, तो OR को $=\frac{a*d}{b*c}$ के रूप में परिभाषित किया जाता है, जहां * गुणन को दर्शाता है।

धूम्रपान और हृदय रोग के आंकड़ों के मामले में हमें OR= (85 * 118) / (28 * 71) = 5.04 मिलता है। यह मान RR से अलग है क्योंकि धूम्रपान करने वालों के साथ हृदय रोग की घटना आम है।

रोग असामान्य या दुर्लभ होने पर OR मान RR के समान होगा। RR नहीं किया जा सकता है, जबकि आगे या एक मामले-नियंत्रण डेटा से गणना की जा सकती है। RR और OR के बारे में अधिक जानकारी इंद्रायन और सत्यनारायण (2006) में देखी जा सकती है।

नीचे चर्चा किए गए कुछ उपायों की मदद से वेरिएबल्स के बीच दो स्पष्ट संबंध की ताकत को समझा जाता है।

यूल का गुणांक (Y): यह 1912 में उडी यूल द्वारा प्रस्तावित दो श्रेणीगत चरों के बीच साहचर्य का एक माप है। Y का मान -1 और +1 के बीच है और सूत्र $Y = \frac{\sqrt{ad}-\sqrt{bc}}{\sqrt{ad}+\sqrt{bc}}$ और Y के बीच सूत्र का उपयोग करके 2 x 2 तालिका से गणना की जाती है - 1 और +1। 1 का मान पूर्ण सकारात्मक साहचर्य को इंगित करता है और -1 का मान पूर्ण नकारात्मक साहचर्य को दर्शाता है।

पियर्सन का फी गुणांक (φ): यह 2x2 आषंग तालिका के काई स्क्वायर मान पर आधारित एक मान है और इसकी गणना सूत्र $\varphi = \sqrt{\chi^2/n} = \mathbf{n}$ का उपयोग करके की जाती है जहां द तालिका में सभी मानों को दर्शाती है। φ का मूल्य -1 और +1 के बीच है।

क्रैमर का वी (Cramer's V: यह 2 या अधिक पंक्तियों या स्तंभों या दोनों वाली आषंग तालिकाओं के लिए लागू साहचर्य का एक मान है। V का मान 0 और 1 के बीच होता है, जैसे 0 कोई साहचर्य नहीं दर्शाता है और 1 पूर्ण साहचर्य को इंगित करता है। सूत्र $V = \sqrt{\frac{\chi^2/n}{\min(k-1, r-1)}}$ है जहां k स्तंभों की संख्या को दर्शाता है और r पंक्तियों की संख्या को दर्शाता है।

अपनी प्रगति जांचें

1) श्रेणीबद्ध चर से क्या अभिप्राय है? क्रमिक और अक्रमिक चर के बीच अंतर लिखिए।

.....

.....

.....

2) रिलेटिव रिस्क एंड ऑड्स रेशियो पर एक संक्षिप्त नोट लिखें।

.....

.....

.....

10.3 रैखिक प्रतिगमन (लीनियर रिग्रेशन)

सरल रैखिक प्रतिगमन एक वैद्यकीय तकनीक है जिसका उपयोग दो चरों के बीच के कारण और प्रभाव संबंध को समझाने के लिए किया जाता है। यह एक गणितीय मॉडल (सूत्र) है, जो कि पूर्वसूचक चर (एक्स) और प्रतिक्रिया चर (वाई) पर अवलोकन के जोड़े के रूप में एकत्र नमूना डेटा से लिया जाता है। यह माना जाता है कि संबंध रैखिक है, इसका मतलब है कि वाई में परिवर्तन एक स्थिर दर पर होता है जिसमें एक्स में एक इकाई परिवर्तन होता है।

प्रतिगमन विश्लेषण आमतौर पर पूर्वानुमान के मानों को पढ़कर प्रतिक्रिया का अनुमान लगाने के लिए उपयोग किया जाता है। ऐसे पूर्वानुमानों को कुछ स्वास्थ्य अध्ययनों में निदान (prognostic) कारक कहा जाता है। वे या तो निरंतर या श्रेणीबद्ध हो सकते हैं। जब एक से अधिक पूर्वानुमान का उपयोग प्रतिक्रिया को समझाने के लिए किया जाता है, तो हम प्रतिगमन को बहुरैखिक प्रतिगमन कहते हैं। प्रतिगमन विश्लेषण की एक और शाखा है जिसे गैर-रेखीय प्रतिगमन कहा जाता है जो वर्तमान इकाई के दायरे से परे है।

प्रतिगमन विश्लेषण सहसंबंध गुणांक की अवधारणा से निकटता से संबंधित है जो उम्र और शरीर के वजन जैसी दो मापा मात्राओं के बीच रैखिक संबंधों की ताकत को मापता है। यह आर द्वारा चिह्नित किया जाता है और नीचे दिए गए पियर्सन के सूत्र का उपयोग करके आकार एन के नमूने से इसकी गणना की जाती है।

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

इस सहसंबंध गुणांक को उत्पाद क्षण सहसंबंध *(product moment correlation)* भी कहा जाता है।

आधुनिक कंप्यूटर त्वरित गणना या r में मदद करते हैं। पेशेवर के लिए रूचि के कुछ बिंदु निम्नलिखित हैं।

1) r का मान -1 और +1 के बीच होता है

2) r = 0 का मतलब कोई रैखिक संबंध नहीं है

3) r का एक छोटा मूल्य कमजोर संबंध को इंगित करता है

4) स्कैटर आरेख का उपयोग करके, संबंध की प्रकृति देखी जा सकती है।

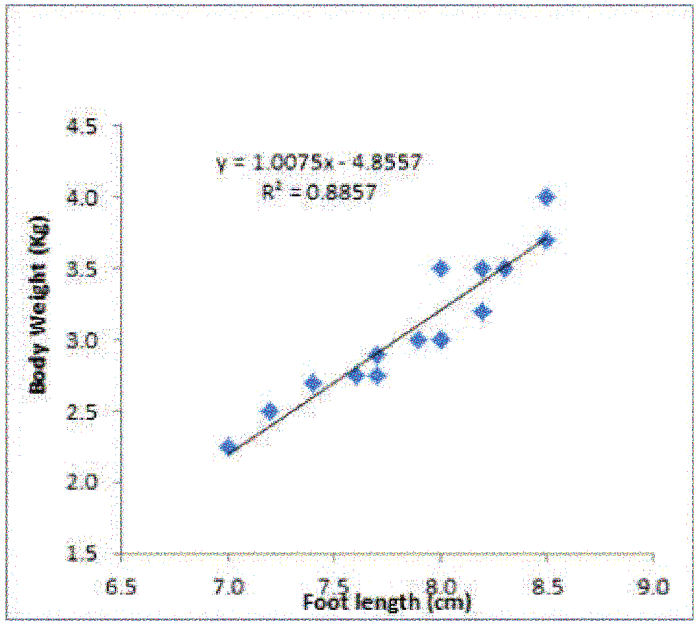
सरल रैखिक प्रतिगमन

सरल रेखीय प्रतिगमन मॉडल $Y = a + bX + e$ द्वारा दिया जाता है जहां 'a' को एक अवरोधन या आधार रेखा मान कहा जाता है और 'इ' को प्रतिगमन गुणांक कहा जाता है। 'e' शब्द को यादृच्छिक त्रुटि घटक कहा जाता है और यह अस्पष्टीकृत कारकों की भूमिका का प्रतिनिधित्व करता है जो एक्स के अलावा वाई को प्रभावित कर सकता है। 'ए' और 'बी' के मूल्यों का अनुमान नमूना डेटा से कम से कम वर्गों की विधि का उपयोग करके लगाया जाता है। यहाँ एक उदाहरण है।

उदाहरण 10.3: 15 नवजात शिशुओं के जन्म का वजन किलो में (वाई) पैर (एक्स) की लंबाई सेमी में के साथ मापा जाता है।

S. No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	7.7	7.9	8.3	8.0	8.2	8.5	7.4	8.0	8.5	7.2	7.0	8.0	7.7	8.2	7.6
Y	2.8	3.0	3.5	3.0	3.2	4.0	2.7	3.5	3.7	2.5	2.3	3.0	2.9	3.5	2.8

हम एक रैखिक प्रतिगमन मॉडल का उपयोग करके जन्म वजन (वजन मशीन के बिना) निर्धारित करने के लिए एक कारक के रूप में पैर की लंबाई का उपयोग करना चाहते हैं। एक्सेल में तैयार स्कैटर आरेख चित्र 10.1 में दिखाया गया है। यह देखा जा सकता है कि पैर की लंबाई और बच्चे के जन्म के वजन के बीच एक रैखिक और सकारात्मक संबंध है।



चित्र 10.1: स्कैटर आरेख और रैखिक प्रतिगमन समीकरण

प्रतिगमन मॉडल को निम्नलिखित चरणों से सुसज्जित किया गया है:

1) स्कैटर चार्ट पर किसी भी डॉट पर राइट क्लिक करें

2) विकल्प का चयन करें 'ट्रेंड लाइन जोड़ें'

3) 'चार्ट पर डिस्प्ले इक्वेशन' विकल्प चुनें

4) 'चार्ट पर डिस्प्ले आर-स्क्वेर्ड वैल्यू' विकल्प चुनें।

5) 'ओके' पर क्लिक करें

इन विकल्पों के साथ हमें फिट के लिए एक माप के साथ चार्ट पर प्रदर्शित प्रतिगमन मॉडल मिलता है, जिसे आर-स्क्वेर मूल्य कहा जाता है। आइए हम परिणामों को समझते हैं:

क) प्रतिगमन मॉडल $Y = -4.856. + 1.008 * X$ है।

ख) गुणांक $= -4.856$ और $b = 1.008$ हैं

ग) एक सेंटीमीटर तक पैर की लंबाई बढ़ाने के लिए, शरीर का वजन 1.008 किलोग्राम बढ़ जाता है। मान लीजिए कि एक बच्चे की लंबाई 8 सेमी है। फिर ऐसे बच्चों के लिए शरीर का अनुमानित वजन 3.20 किलोग्राम होगा।

घ) $R2 = 0.8857$ का मान जो बताता है कि शरीर के वजन का 88.57% इस मॉडल का उपयोग करके पैर की लंबाई से समझाया जा सकता है।

अं) जन्म के वजन और पैर की लंबाई के बीच सहसंबंध गुणांक 0.8857 का वर्गमूल है। इसलिए $r = 0.9411$ जो यह दर्शाता है कि पैर की लंबाई जन्म के वजन का एक अच्छा पूर्वानुमान है।

च) सहसंबंध गुणांक का चिन्ह प्रतिगमन गुणांक 'b' (+1.008) के समान है और इसलिए इस मामले में सहसंबंध सकारात्मक है।

हाथ की गणना के साथ काम करते समय हम निम्नलिखित सूत्रों का उपयोग करते हैं।

$$b = \frac{\sum(X - \overline{X})(Y - \overline{Y})}{\sum(X - \overline{X})^2} \text{ and } a = \overline{Y} - b\overline{X}$$

जहाँ $\overline{X}$ और $\overline{Y}$ क्रमशः X और Y का माध्य.दर्शाते हैं

बहु रेखीय प्रतिगमन

बहु रैखिक प्रतिगमन साधारण प्रतिगमन का एक विस्तार है। यह मॉडल का उपयोग करके वाई को एक से अधिक व्याख्यात्मक चर X1, X2,..., Xk से संबंधित करता है

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + e$$

जहाँ $\beta_1, \beta_2, ..., \beta_k$ व्याख्यात्मक चर के प्रतिगमन गुणांक (भार) को निरूपित करें और β_0 एक स्थिर है। β_0 का मान Y के औसत का प्रतिनिधित्व करता है जब सभी X चर शून्य पर सेट होते हैं। शब्द 'ई' यादृच्छिक त्रुटि घटक का प्रतिनिधित्व करता है। गुणांक β_1, $\beta_2, ..., \beta_k$ न्यूनतम वर्गों की विधि से नमूना डेटा से अनुमानित हैं। प्रक्रिया में जटिल गणना शामिल है और एक अच्छी विधि एक्सेल या एसपीएसएस जैसे सॉफ्टवेयर का उपयोग करना है। एक बार जब हम गुणांक का अनुमान लगाते हैं, तो हम कहते हैं कि मॉडल डेटा के लिए फिट है।

समंजन-सुष्ठुता मॉडल (goodness of the fitted model) को R2 नामक एक माप द्वारा आंका जाता है जो 0 और 1 के बीच स्थित है। उच्च मूल्य बेहतर फिट को इंगित करता है। चूँकि मॉडल की फिटिंग नमूना डेटा पर आधारित है, इसलिए यह आवश्यक है कि (i) प्रत्येक प्रतिगमन गुणांक (छात्र के टी-टेस्ट द्वारा) और (ii) R2 मान (एफ-टेस्ट द्वारा) के सांख्यिकीय महत्व के लिए परीक्षण किया जाए। कन्वेंशन पी-वैल्यू द्वारा <0-05 महत्व को इंगित करता है। सरमा (2010) ने कई प्रतिगमन पर विवरणों पर चर्चा की जिसमें एसपीएसएस का उपयोग किया जा सकता है।

यहाँ एक उदाहरण है।

उदाहरण 10.4: निम्नलिखित डेटा का तात्पर्य उन 100 बिंदुओं पर मापे गए जीवन की गुणवत्ता (क्वालिटी ऑफ लाइफ) से है, जो हृदय की सर्जरी से गुजर चुके हैं। इन मरीजों को फिजियो-थेरेपी की सलाह दी गई। क्यूओएल आयु, लिंग, फिजियोथेरेपी की अवधि (दिन) और अस्पताल में रहने की अवधि (दिन) पर निर्भर पाया जाता है। चर को X1 = आयु (वर्ष), X2 = लिंग (पुरुष = 1, महिला = 2), X3 = भौतिक चिकित्सा की अवधि (दिन), X4 = अस्पताल में रहने की अवधि (दिन) और Y = Q स्कोर (अधिकतम= 100) के रूप में कोडित किया गया है। 25 रिकॉर्ड के साथ नमूना डेटा नीचे दिया गया है।

तालिका 10.2: 25 मरीजों में जीवन की गुणवत्ता का डेरा

S. No	X_1	X_2	X_3	X_4	Y
1	21	1	12	5	65
2	25	1	10	8	58
3	26	2	6	5	59
4	26	2	10	3	63
5	26	1	12	4	64
6	27	1	6	4	61
7	28	2	11	3	65
8	28	1	10	2	67
9	29	1	11	8	54
10	30	1	12	6	62
11	31	1	8	7	60
12	31	1	6	6	59
13	32	2	13	5	65
14	32	2	7	7	57
15	32	1	14	2	65
16	34	2	10	4	65
17	34	2	10	3	63
18	35	2	5	8	54
19	35	1	8	5	63
20	36	1	15	5	66
21	36	2	9	7	56
22	36	2	15	4	67
23	38	2	10	3	68
24	39	1	8	8	61
25	40	1	11	3	63

हम Y से X_1, X_2, X_3 और X_4 से संबंधित कई रैखिक प्रतिगमन मॉडल बनाना चाहते हैं।

विकल्पों के साथ एसपीएसएस का उपयोग करना Analyze → Regression → Linear हमें उपयुक्त इनपुट बॉक्स में निर्भर चर और अन्य स्वतंत्र चर का चयन करने के लिए विकल्प विंडो मिलती है। आइए हम 'विधि' को 'स्टेपवाइज' के रूप में चुनें और ओके दबाएं। यह आउटपुट देता है जिसे नीचे संक्षेप में प्रस्तुत किया गया है।

मॉडल में R2 = 0.75 है जिसका अर्थ है कि वाई के व्यवहार का 75% मॉडल द्वारा समझाया जा सकता है।

1) F-test (ANOVA तालिका में दिया गया) उच्च महत्व (p <0-001) दर्शाता है जिसका अर्थ है कि मॉडल की अच्छाई 'संयोग से घटना नहीं है '।

2) प्रतिगमन गुणांक (व्याख्यात्मक चर के भार) निम्नानुसार हैं।

तालिका 10.3: Output of Multiple Linear Regression

Variable	*Coefficients*	*Standard Error*	*t Stat*	*P-value*
Intercept	62.202	3.944	15.77	<0.001
Age	0.075	0.093	0.800	0.433
Gender	-0.506	0.919	-0.551	0.5871
Duration of Physiotherapy	0.500	0.177	2.811	0.011*
Duration of Hospital Stay	-1.365	0.252	-5.398	< 0.001*
* Regression coefficient is statistically significant.				

यह देखा जा सकता है कि फिजियोथेरेपी की अवधि और अस्पताल में रहने की अवधि क्यूओएल पर महत्वपूर्ण प्रभाव डालती है। (पी-मान * के साथ चिह्नित)। इंटरसेप्ट (मॉडल का निरंतर घटक) भी महत्वपूर्ण है लेकिन हम ज्यादातर समय पूर्वानुमानों के महत्व में रुचि रखते हैं।

हम इस खंड को इस अवलोकन के साथ समाप्त करते हैं कि रैखिक प्रतिगमन अध्ययन के परिणामों पर व्याख्यात्मक चर के प्रभाव को समझाने के लिए सांख्यिकीय मॉडल का प्रस्ताव करने के लिए एक उपयोगी उपकरण है। एसपीएसएस के साथ प्रतिगमन विश्लेषण के संचालन के बारे में अधिक विवरण सरमा (2010) में देखा जा सकता है।

10.4 प्रसरण विश्लेषण (ANOVA) एनोवा

प्रसरण विश्लेषण (ANOVA) एक सांख्यिकीय उपकरण है जिसका उपयोग तीन या अधिक स्वतंत्र समूहों के बीच एक एकल निरंतर चर (Y) के औसत मूल्यों की तुलना करने के लिए किया जाता है। समूहीकरण चर को आयु वर्ग की खुराक, सामाजिक आर्थिक स्थिति आदि; जिसे कारक कहा जाता है, जो कुछ स्तरों के साथ स्पष्ट है। एनोवा परीक्षण में मदद करता है कि क्या समूह का माध्य काफी अलग है। इसे टू-सैंपल टी-टेस्ट के विस्तार के रूप में माना जा सकता है। यदि वाई को प्रभावित करने वाला केवल एक कारक है, तो हम वन-वे एनोवा का उपयोग करते हैं लेकिन सामान्य तौर पर हमारे पास एक से अधिक कारक हो सकते हैं और हम सभी कारकों और उनके संयोजन के कारण महत्व का परीक्षण कर सकते हैं। मान लीजिए कि हम विभिन्न तरीकों से वजन घटाने के लिए उपचार प्राप्त करने वाले व्यक्तियों के समूह की प्रतिक्रिया वाई (बॉडी मास इंडेक्स की तरह) को माप रहे हैं। ए (फूड कंट्रोल), बी (व्यायाम) और सी (दोनों खाद्य नियंत्रण और व्यायाम)। हम यह जांचना चाहते हैं कि तीनों समूहों में Y का माध्य समान है या नहीं। चूंकि डेटा को केवल एक कारक के अनुसार वर्गीकृत किया जाता है, इसलिए इसे वन-वे एनोवा कहा जाता है।

A , B और C के समूहों में Y के साधनों को क्रमशः $\mu 1, \mu 2$ और $\mu 3$ द्वारा निरूपित करें। फिर हम अशक्त परिकल्पना H_0: $\mu 1 = \mu 2 = \mu 3$ का परीक्षण करना चाहते हैं। वैकल्पिक परिकल्पना H1: कम से कम दो साधन समान नहीं हैं। यह माना जाता है कि वाई का विचरण प्रत्येक समूह में समान रहता है। इसका मतलब है कि प्रतिक्रिया सभी समूहों में सुसंगत है और इसमें बहुत अधिक या निम्न मान नहीं हैं।

फिर शून्य परिकल्पना का परीक्षण एफ-अनुपात नामक अनुपात का उपयोग करके किया जाता है

$$F = \frac{\text{Mean Sum of Squares (MSS) due to the factor}}{\text{Residual MSS}}$$

एफ-अनुपात परीक्षण मूल्य है। जब तीनों का मतलब एक बड़े अंतर से अलग होता है तो हम एक उच्च एफ-मूल्य रखते हैं। यदि पी-मान 0.05 से कम है तो हम शून्य परिकल्पना को अस्वीकार कर सकते हैं और विचार कर सकते हैं कि कारक का वाई पर महत्वपूर्ण प्रभाव है, अन्यथा शून्य परिकल्पना को स्वीकार करें। जब एफ-अनुपात महत्वपूर्ण होता है, तो हम अनुमान लगाते हैं कि समूह मानों में अंतर संयोग की घटना नहीं है।

जब शून्य परिकल्पना को खारिज कर दिया जाता है, तो हमें यह पहचानना होगा कि कारक के किन स्तरों पर मान अलग-अलग हुए हैं। यह कई तुलना परीक्षण या युग्मक तुलना परीक्षण का उपयोग करके किया जाता है। डंकन के मल्टीपल रेंज टेस्ट (DMRT), कम से कम मानक विचलन (LSD) परीक्षण या शेफी के परीक्षण जैसे कई परीक्षण हैं। एनोवा के लिए गणना या तो एमएस-एक्सेल के साथ या एसपीएसएस के साथ की जा सकती है।

उदाहरण 10.5: एक शोधकर्ता ने हार्मोन थेरेपी से गुजरने के बाद रोगियों के बॉडी मास इंडेक्स (बीएमआई) को मापा है। रोगियों की आयु को तीन आयु समूहों में वर्गीकृत किया गया था: i) <30 वर्ष, ii) 31– 40 वर्ष और iii) 41 और उससे अधिक क्रमशः 1, 2, 3। डेटा में BMD (हड्डी खनिज घनत्व) नामक एक अन्य प्रतिक्रिया चर भी है जैसा कि तालिका 10.4 में दिखाया गया है। हम यह जांचना चाहते हैं कि क्या औसत आयु वर्ग के बीच बीएमआई रहता है।

विश्लेषण:

ANOVA की गणना एक्सेल या SPSS के साथ आसानी से की जा सकती है। हम एसपीएसएस के साथ इसका उदाहरण देते हैं। जिसमें निम्नलिखित विकल्प हैं।

1) SPSS डेटा फाइल खोलें
2) Analyze → Compare Means →One way ANOVA चुनें
3) variable निर्भर चर 'बॉक्स में बीएमआई का चयन करें
4) 'फैक्टर' विंडो में आयु समूह का चयन करें
5) विकल्प टैब पर क्लिक करें और वर्णनात्मक आंकड़े चुनें
6) ओके दबाएं।

तालिका 10.4: BMI and BMD data

S. No	BMI	Age group	BMD
1	21.5	2	0.933
2	22.0	2	0.889
3	22.8	2	0.937
4	22.7	3	0.874
5	23.1	2	0.953
6	22.9	2	0.671
7	23.1	3	0.914
8	18.3	1	0.883
9	22.9	2	0.749
10	18.0	1	0.875
11	22.1	3	0.715
12	23.8	3	0.932
13	23.5	2	0.800
14	23.8	3	0.699
15	22.1	3	0.677
16	20.8	2	0.813
17	18.0	1	0.851
18	19.2	1	0.888
19	17.8	1	0.875
20	20.1	1	0.773

हम पहले प्रत्येक समूह में बीएमआई के माध्य और मानक विचलन की रिपोर्ट करेंगे जैसा कि नीचे दिखाया गया है। (एसपीएसएस आउटपुट वास्तव में माध्य और मानक विचलन के अलावा मानक त्रुटि और 95% आत्मविश्वास अंतराल दिखाता है)।

Age group	N	Mean	Std. Deviation
< 30	6	18.56	0.900
31 – 40	8	22.43	0.916
41 & above	6	22.93	0.771
Total	20	21.42	2.099

ANOVA इसके मानक प्रारूप में तालिका 10.5 दिखाया गया है। रिपोर्ट लिखने के लिए हम एफ-मूल्य और संबंधित पी-मूल्य (एसपीएसएस आउटपुट में Sig- द्वारा इंगित) प्रस्तुत कर सकते हैं। हम निम्नलिखित घटकों को समझते हैं।

1) भिन्नता के दो स्रोत हैं, 'समूहों के बीच' (आयु समूह के कारण) और 'समूहों के भीतर' (यादृच्छिक और अनियंत्रित कारकों का संकेत जो बीएमआई को प्रभावित कर सकते हैं)। बीएमआई में कुल भिन्नता का योग हैः ए) अज्ञात कारकों के कारण (आयु समूह) और बी) ज्ञात कारक के कारण है।

तालिका 10.5: एक तरफा एनोवा तालिका

Source of variation	Sum of squares	d. f.	Mean Square	F	Sig.
Between Groups	70.872	2	35.436	46.679	0.0001
Within Groups	12.905	17	0.759		
Total	83.778	19			

2) स्वतंत्रता का अंश (df) मीन स्क्वायर में प्रयोग किए जाने वाले विभाजक का सूचक है। Df स्वतंत्र प्रेक्षणों (साधनों) की संख्या को इंगित करता है और सामान्य सूत्र df = (k–1) है यदि k— अवलोकन हैं। इसीलिए कुल df (20-1)= 19 है। चूंकि 3 समूह हैं जो हमें मिलते हैं (3-1) = 2 df अंत में, समूहों के घटक के लिए df 19 है (घटाव के द्वारा)।

3) वर्गों का योग और वर्गों का औसत योग मध्यवर्ती गणनाएं हैं, जो हमें आयु वर्ग के कारण अनुमानित विचरण का पता लगाने के लिए प्रेरित करती हैं। एफ-वैल्यू को एफ-अनुपात या विचरण अनुपात कहा जाता है। शीर्षक 'sig' एफ-अनुपात के पी-मूल्य को इंगित करता है। चूंकि पी-मान <0-05 है, हम शून्य परिकल्पना को खारिज करते हैं और निष्कर्ष निकालते हैं कि तीन आयु समूहों के बीच बीएमआई का माध्य काफी भिन्न होता है।

4) शास्त्रीय दृष्टिकोण में, सांख्यिकीय तालिकाओं से प्राप्त एफ-अनुपात के क्रांतिक मान का उपयोग किया जाता है। इस मामले में स्वतंत्रता के 5% के स्तर पर (2.17) डिग्री के लिए एफ-क्रिटिकल मूल्य 3.59 है। चूंकि प्राप्त मूल्य इस से अधिक है, इसलिए हम शून्य परिकल्पना को अस्वीकार करते हैं।

समूह की तुलनात्मक रूप से तुलना का मतलब है (जैसे 1 बनाम 2, 1 बनाम 3 और 2 बनाम 3) डंकन के परीक्षण के साथ नीचे दिखाए गए पोस्ट-हॉक प्रक्रिया के रूप में किया जाता है। हमें यहां टू-सेंपल और टी-टेस्ट का उपयोग नहीं करना चाहिए!

डंकन का परीक्षण एनोवा तालिका में गणना किए गए मीन योगों का उपयोग करता है। इसमें 2 सबसेट हैं जिनमें तीन मानों को वर्गीकृत किया गया है।

Duncan's test for comparing the mean BMI among age groups			
Age group	N	Subset for alpha = 0.05*	
		1	2
< 30	6	18.56	---
31 – 40	8	---	22.43
41 & above	6	---	22.93
Sig.		1.000	0.318
* Means for groups in homogeneous subsets are displayed.			

वे अर्थ जो एक ही सबसेट के हैं उन्हें समरूप (पी-मान देखें) के रूप में माना जाता है, इस अर्थ में, वे काफी भिन्न नहीं हैं। आयु समूहों '31 '-30' और '41 और उससे अधिक' में औसत बीएमआई का कोई महत्वपूर्ण अंतर नहीं है (पी = 0.318) लेकिन दोनों का मान बीएमआई से अलग है '<30' आयु वर्ग।

यह वन वे एनोवा को पूरा करता है।

अपनी प्रगति जाँचें

3) बहु रैखिक प्रतिगमन क्या है? इसकी गणना एस.पी.एस.एस द्वारा कैसे की जाति है?

.....

.....

.....

4) एनोवा (ANOVA) परीक्षण क्या है? इसकी गणना एस.पी.एस.एस द्वारा कैसे की जाती है?

.....

.....

.....

टिप्पणी

क) बार चार्ट द्वारा माध्य को प्रदर्शित करना भी एक अभ्यास है, लेकिन एसपीएसएस लाइन चार्ट द्वारा इसे प्रदर्शित करता है ।

ख) आयु समूह के बजाय, यदि हम एसपीएसएस विकल्पों में कारक के रूप में वास्तविक आयु (वर्ष) का उपयोग करते हैं, तो हमें एक अप्रिय आउटपुट मिलता है! केवल श्रेणीबद्ध चर (अक्रमिक या क्रमिक पैमाने पर) का उपयोग किया जाएगा। ऐसा नहीं करना चाहिए।

ग) यहाँ उपयोग किए गए एनोवा को एक चरीय एनोवा कहा जाता है क्योंकि केवल एक प्रतिक्रिया चर को कई समूहों के बीच माना जाता है। यदि एक समय में दो या अधिक प्रतिक्रियाओं का अध्ययन किया जाता है, तो हम इसे एक प्रोफाइल कहते हैं जिसे हमें एक अग्रिम उपकरण का उपयोग करना होगा जिसे मल्टीवेरेट एनोवा या मैनोवा कहा जाता है।

घ) एक से अधिक कारक वाले एनोवा एसपीएएस में विश्लेषण विकल्प में उपलब्ध सामान्य रैखिक मॉडल का उपयोग करके एसपीएसएस द्वारा किया जाता है।

हम इस चर्चा को इस अवलोकन के साथ समाप्त करते हैं कि एनोवा विधि एक सांख्यिकीय निष्कर्ष है और इसमें सावधानीपूर्वक व्याख्या की आवश्यकता है। केवल पी-वैल्यू की रिपोर्ट करना पर्याप्त नहीं है। हमें इस पर टिप्पणी करनी होगी कि समूहों में माध्य मान कैसे भिन्न होते हैं।

10.5 सारांश

- हमने सीखा है कि श्रेणीबद्ध चर के बीच संगति का मापन काई स्क्वायर परीक्षण द्व ारा किया जाता है जो एक आषंग तालिका पर आधारित है। कुछ मानक उपायों में यूल के Y, पियर्सन की फी φ और क्रैमर की V सांख्यिकी शामिल हैं। मात्रात्मक डेटा (एक अंतराल के पैमाने पर मापा गया) के मामले में हम पियर्सन के सहसंबंध गुणांक का उपयोग करते हैं।

- हमने यह भी देखा है कि सहसंबंध विश्लेषण के विपरीत प्रतिगमन विश्लेषण (एनोवा) चर के बीच संबंध के रूप को मापता है। चरणबद्ध प्रतिगमन एक कार्यात्मक प्रतिगमन मॉडल स्थापित करने के लिए एक अनुशंसित विधि है।

- आगे हमने डेटा सेट के तीन या अधिक समूहों के बीच एक विशेषता के औसत मूल्यों की तुलना करने के लिए एनोवा के सिद्धांत को समझा है। यह एफ-परीक्षण पर आधारित है और जिसकी एसपीएसएस के साथ गणना की जा सकती है। विश्लेषण केवल तभी पूरा होता है जब कई तुलना परीक्षण (जैसे डंकन का परीक्षण) किया जाता है।

10.6 संदर्भ

इंद्रायन, ए., एंड सत्यनारायण, एल.(2006). *बायोस्टैटिस्टिक्स फॉर मेडिकल, नर्सिंग एंज फार्मेसी स्टूडेंट.* नई दिल्लीः प्रेंटिस हॉल ऑफ इंडिया.

राव, पी.एस.,एंड रिचर्ड, जे.(2012). *इंट्रोडक्शन टू बायोस्टैटिस्टिक्स एंड रिसर्च मैथड्स,* 5वां संस्करण, नई दिल्लीः प्रेंटिस हॉल ऑफ इंडिया.

सरमा, के.वी. एस. (2010). *स्टटिस्टिक्स मेड सिंपल डू इट योरसेल्फ ऑन पीसी,* दूसरा संस्करण.नई दिल्लीः प्रेंटिस हॉल ऑफ इंडिया.

10.7 आपकी प्रगति की जाँच करने के लिए उत्तर

1) श्रेणीबद्ध चर को गुणात्मक कारकों के रूप में भी जाना जाता है। ऐसे कारकों पर डेटा एक माप नहीं है, लेकिन यह संभावित मूल्यों की असतत सूची से चुनने के लिए एक विकल्प की तरह है। विवरण के लिए भाग 10.0 देखें।

2) महामारी विज्ञान के कुछ अध्ययनों में हम सापेक्ष जोखिम और बाधाओं जैसे उपायों के बारे में बताते हैं, जो दोनों की गणना के 2 x 2 तालिका पर आधारित हैं। एक आबादी के लोगों के अनुपात, उन सभी के बीच बीमारी होना जो किसी स्थिति के संपर्क में हैं, को बीमारी का जोखिम कहा जाता है जबकि ऑड अनुपात (OR) केस-कंट्रोल अध्ययनों के संदर्भ में उपयोग किए गए संबंधों (एसोसिएशन) का एक और उपाय है। विवरण के लिए भाग 10.2 देखें।

3) एकाधिक रैखिक प्रतिगमन साधारण प्रतिगमन का एक विस्तार है। यह एक से अधिक व्याख्यात्मक चर के साथ वाई से संबंधित है। अधिक विवरण के लिए भाग 10.3 देखें।

4) प्रसरण विश्लेषण (ANOVA) एक सांख्यिकीय उपकरण है जिसका उपयोग तीन या अधिक स्वतंत्र समूहों के बीच एक एकल चर (वाई) के औसत मूल्यों की तुलना करने के लिए किया जाता है। समूहीकरण चर को आयु वर्ग की खुराक, सामाजिक आर्थिक स्थिति आदि; जिन्हें कारक कहा जाता है, जो कुछ स्तरों के साथ स्पष्ट होते है। विवरण के लिए भाग 10.4 देखें।

सुझावित अध्ययन

खंड 1: महामारी विज्ञान और सार्वजनिक स्वास्थ्य में अनिवार्य तत्व

बीगलहोल, आर. एंड बोनिता, आर.(1997). *पब्लिक हेल्थ एट द क्रॉसरोडः अचीवमेंट एंड प्रास्पैक्ट.* कैम्ब्रिजः कैम्ब्रिज यूनिवर्सिटी प्रेस.

ब्लूमेंथल, डीएस एंड रुटेनबेर, ए.ए.ए(1995). *इंट्रोडक्शन टू इनवायमेंटल हेल्थ.* दूसरा संस्करण. न्यूयॉकः स्प्रिंगर.

लास्ट, जॉन एम. (1998). *पब्लिक हेल्थ एंड ह्युमन इकोलॉजी.* लंदनः प्रेंटिस हॉल.

श्नाइडर, मैरी– जेन. (2006). *इंट्रोडक्शन टू पब्लिक हेल्थ.* लंदनः जोन्स और बार्टलेट.

टर्नकॉक, बी.(1994). *पब्लिक हेल्थ.* बोस्टनः जोन्स और बार्टलेट.

पार्क, के. (2007). *पार्क्स टेक्स्ट बुक ऑफ प्रिवेंटिव एंड सोशल मेडिसिन.* जबलपुरः बनारसीदास भनोट पब्लिशर्स.

ग्रोवर, ए. एंड सिंह आर.बी. (2019). *अर्बन हेल्थ एंड वेलबिंग.* जापानः स्प्रिंगर

महाजन, एम.सी., गुप्ता बी.के.(2013). *टैक्स्ट् बुक ऑफ प्रिवेंटिव एंड सोशल मेडिसिन.* चौथा संस्करण, आर.एन. रॉय आई. साहा द्वारा संशोधित, जेपी ब्रदर्स मेडिकल पब्लिशर्स लि.

https://mohfw.gov.in/

https://www.undp.org/content/undp/en/home/sustainable-development-goals.html

खंड 2: सार्वजनिक स्वास्थ्य और प्रबंधन में मनोवैज्ञानिक, व्यवहारिक और सामाजिक मुद्दे

गुप्ता मोनिका (2016). *पब्लिक हेल्थ इन इंडियाः* एन ओवरव्यू . वर्किंग पेपर सीरीज 3787. वर्ल्ड हेल्थ आर्गेनाइजेशन.

ग्लान्ज, के., रिमेर, बी.एंड विश्वनाथ, के. (संपा.) (2008). *हेल्थ बिहेवियर एंड हेल्थ एजुकेशन* थियरी, रिसर्च एंड प्रैक्टिस .सैन फ्रैंसिस्कोः विली इंप्रिट.

जेनकिंस, डेविड. (2003). *बिल्डिंग बैटर हेल्थः ए हैंडबुक ऑफ बिहेवियरल चेंज.* वाशिंगटन डीसीः पैन अमेरिकन हेल्थ आर्गेनाइजेशन.

कावाची, आई., एंड वमाला, एस. (संपा.).(2006). *ग्लोब्लाइजेशन एंड हेल्थ.* ऑक्सफोर्ड यूनिवर्सिटी प्रेस.

क्लेनमैन, ए. एंड बेन्सन, पी.(2006). *एंथ्रोपोलॉजी इन द क्लिनिक PLoS मेडिसिन* (10): 294.

क्लेनमैन, ए. (2004). *कल्चर एंड साइकैट्रिक डायग्नोसिस एंड ट्रिटमेंटः* ट्रिम्बोस लेक्चर. हार्वर्ड यूनिवर्सिटी.

पार्क, के.(2007). *पार्क्स टेक्स्ट बुक ऑफ प्रिवेंटिव एंड सोशल मेडिसिन.* जबलपुरः बनारसीदास भनोट पब्लिशर्स.

राहेल डेविस, रोना कैंपबेल, जो हिल्डन, लोर्ना हॉब्स एंड सुसान मिसी (2015). *थियरी*

ऑफ बिहेवियर एंड बिहेवियरल चेंज एक्रास द सोशल एंड बिहेवियरल साइंसेज : एक स्कूपिंग रिव्यू. हेल्थ साइकालजी रिव्यू, 9: 3, 323–344.

वर्ल्ड हेल्थ आर्गेनाइजेशन (2008).जेनेवा: *कमीशन ऑन सोशल डिटर्मिनेंट्स ऑफ हेल्थ रिपोर्ट.*

खंड 3: सार्वजनिक स्वास्थ्य में अनुसंधान और सांख्यिकीय विधियां

इंद्रायन,ए.एंड सत्यनारायना,एल.(2006). *बायोस्टटिस्टिक्स फॉर मेडिकल, नर्सिंग एंड फार्मेसी स्टूडेंट्स.* न्यू दिल्ली : प्रेंटिस हॉल ऑफ इंडिया.

माइक्रोसॉफ्ट एक्सल (2010) स्टेप बाय स्टेप (ई-बुक) वेब स्रोत: https://www.spss-tutorials.com/basics/

सबाइन लैंडौ एंड ब्रायन एस. ई. (2004). *ए हैंडबुक ऑफ स्टटिस्टिकल एनलॉसिस यूसिंग एसपीएसएस.* यूएसए: चौपमैन एंड हॉल/सीआरसी प्रेस एलएलसी.

सुंदर लाल एंड विकास (2018). *पब्लिक हेल्थ मैनेजमेंट – प्रिसिंपल एंड प्रैक्टिस,* दिल्ली: सीबीएस पब्लिशर एंड डिस्ट्रिब्यूटर्स प्राइवेट लिमिटेड.

सुरेश के शर्मा (2014). *नर्सिंग रिसर्च एंड स्टटिस्टिक* (दूसरा संस्करण). गुरुग्राम: एल्सेवियर आरईएलएक्स इंडिया प्राइवेट लिमिटेड.

वेन डब्ल्यू डैनियल (2014). *बायोस्टैटिस्टिक्स: ए फाउंडेशन फॉर एनलॉसिस इन द हेल्थ साइंसेज.* विले सिरीज इन प्रोबेबिलेटी एंड स्टटिस्टिक्स.