अध्याय

2

11099CH02

आँकड़ों का संग्रह

> ***इस अध्याय के अध्ययन के बाद आप इस योग्य होंगे कि:***
> - *आँकड़ा-संग्रह का अर्थ और उद्देश्य समझ सकें;*
> - *प्राथमिक एवं द्वितीयक स्रोतों के बीच अंतर कर सकें;*
> - *आँकड़ा-संग्रह की विधि समझ सकें;*
> - *जनगणना एवं प्रतिदर्श सर्वेक्षण के बीच अंतर कर सकें;*
> - *प्रतिचयन की प्रविधि से परिचित हो सकें;*
> - *द्वितीयक आँकड़ों के कुछ महत्वपूर्ण स्रोतों के बारे में जान सकें।*

प्रस्तावना

पिछले अध्याय में आपको अर्थशास्त्र की विषयवस्तु की जानकारी मिली। इसके साथ ही आपने अर्थशास्त्र में सांख्यिकी की भूमिका एवं महत्त्व के बारे में भी पढ़ा। इस अध्याय में आप आँकड़ों के स्रोतों एवं आँकड़ा-संग्रह की विधि के बारे में अध्ययन करेंगे। आँकड़ों के संग्रह का उद्देश्य किसी समस्या के स्पष्ट एवं ठोस समाधान के लिए साक्ष्य को दर्शाना है।

अर्थशास्त्र में प्राय: ऐसे कथनों से आपका सामना होता है, जैसे–

"अनेक उतार-चढ़ावों के पश्चात् खाद्यान्नों का उत्पादन 1970–71 में 10.8 करोड़ टन से बढ़कर 1978–79 में 13.2 करोड़ टन हो गया, किंतु 1979–80 में फिर से गिर कर 10.8 करोड़ टन हो गया। उसके बाद खाद्यान्नों का उत्पादन 2015–16 तक लगातार बढ़ कर 25.2 करोड़ टन हो गया तथा 2016–17 में इसने 27.2 करोड़ टन का आँकड़ा छू लिया।"

आप इस कथन में यह देख सकते हैं कि विभिन्न वर्षों में खाद्यान्नों का उत्पादन एक समान नहीं रहा है। यह फसल-दर-फसल तथा वर्ष-दर-वर्ष बदलता रहा है। चूँकि ये मूल्य परिवर्तनशील होते हैं, अत: इन्हें चर कहा जाता है। इन चरों को प्राय: X, Y, Z आदि अक्षरों द्वारा प्रदर्शित किया जाता है। प्रत्येक चर का

© NCERT not to be republished

मूल्य प्रेक्षण कहलाता है। उदाहरण के लिए - भारत में खाद्यान्न उत्पादन 1970-71 में 108 मिलियन टन से लेकर वर्ष 2016-17 में 272 मिलियन टन के बीच रहा, जैसा कि सारणी में दिखाया गया है। यहाँ पर वर्षों को चर X के द्वारा और भारत में खाद्यान्नों के उत्पादन को (मिलियन टनों में) चर Y के द्वारा प्रस्तुत किया गया है:

सारणी 2.1
भारत में खाद्यान्नों का उत्पादन
(मिलियन टन में)

X	Y
1970–71	108
1978–79	132
1990–91	176
1997–98	194
2001–02	212
2015–16	252
2016–17	272

यहाँ पर चर X तथा Y के मूल्य 'आँकड़े' हैं, जिनके द्वारा हम भारत में खाद्यान्नों के उत्पादन के बारे में जानकारी प्राप्त कर सकते हैं। खाद्यान्नों के उत्पादन में उतार-चढ़ाव की प्रवृत्ति को जानने के लिए हमें विभिन्न वर्षों के लिए भारत में खाद्यान्न उत्पादन के 'आँकड़ों' की आवश्यकता पड़ती है। आँकड़ा एक ऐसा साधन है, जो सूचनाएँ प्रदान कर समस्या को समझने में सहायक होता है।

आप जानना चाहते होंगे कि ये 'आँकड़े' कहाँ से आते हैं और हम इन्हें कैसे संगृहीत करते हैं? निम्नलिखित अनुभाग में हम आँकड़ों के प्रकार, आँकड़ों को संगृहीत करने की विधि तथा साधनों तथा आँकड़ों के स्रोतों की चर्चा करेंगे।

2. आँकड़ों के स्रोत क्या हैं?

सांख्यकीय आँकड़े दो स्रोतों से प्राप्त किए जा सकते हैं। *गणनाकार* (वह व्यक्ति जो आँकड़ा संग्रह करता है) जाँच-पड़ताल या पूछताछ करके आँकड़े एकत्र कर सकता है। ऐसे आँकड़े *प्राथमिक आँकड़े* कहे जाते हैं, चूँकि ये प्रत्यक्ष रूप से प्राप्त की गई जानकारी पर आधारित होते हैं। मान लें कि आप विद्यालयी बच्चों के बीच किसी फिल्मी सितारे की लोकप्रियता की जानकारी लेना चाहते हैं। इस संबंध में वांछित जानकारी लेने के लिए आपको काफी बड़ी संख्या में विद्यालय के छात्रों से प्रश्नों के माध्यम से पूछताछ करनी होगी। इस विधि से आप जो आँकड़े प्राप्त करते हैं, वह प्राथमिक आँकड़ों का एक उदाहरण है।

यदि किसी दूसरी संस्था द्वारा इन आँकड़ों को संगृहीत एवं संशोधित (संवीक्षित एवं सारणीकृत) किया जाता है तो इन्हें 'द्वितीयक आँकड़े' कहते हैं। इन आँकड़ों को या तो प्रकाशित स्रोतों से जैसे सरकारी रिपोर्ट, दस्तावेज, समाचार पत्र, अर्थशास्त्रियों द्वारा लिखित पुस्तकें, या किसी अन्य स्रोत से प्राप्त किया जा सकता है, जैसे वेबसाइट। अतः ये आँकड़े उन स्रोतों के लिए *प्राथमिक* हैं जो उन्हें पहली बार संगृहीत एवं संसाधित करते हैं, तथा बाद में प्रयोग करने वाले सभी स्रोतों के लिए ये *द्वितीयक* हैं। *द्वितीयक आँकड़ों* के उपयोग से समय एवं धन की बचत होती है। उदाहरण के लिए, मान लें छात्रों में किसी सिनेमा कलाकार की लोकप्रियता के बारे में आँकड़ों को एकत्र करने के पश्चात् आप एक रिपोर्ट प्रकाशित करते हैं। यदि आप द्वारा संग्रह किए गए आँकड़ों का उपयोग कोई इसी तरह के किसी अध्ययन के लिए करता है, तो उसके लिए यह द्वितीयक आँकड़े हो जाते हैं।

3. हम आँकड़े कैसे संगृहीत करते हैं?

क्या आप जानते हैं कि कोई विनिर्माता अपने किसी उत्पाद के संबंध में या कोई राजनैतिक पार्टी अपने किसी उम्मीदवार के विषय में कैसे निर्णय करती है? वे उत्पाद-विशेष या उम्मीदवार-विशेष के बारे में जन- समुदाय से प्रश्नों के माध्यम से सर्वेक्षण करते

© NCERT not to be republished

हैं। इस सर्वेक्षण का उद्देश्य कुछ विशिष्टताओं जैसे कीमत, गुणवत्ता, उपयोगिता (उत्पाद के संबंध में) और लोकप्रियता, ईमानदारी और निष्ठा (उम्मीदवार के संबंध में) के बारे में जानकारी एकत्र करना होता है। सर्वेक्षण का उद्देश्य आँकड़ों को संगृहीत करना होता है। सर्वेक्षण वह विधि है, जिसके द्वारा विभिन्न व्यक्तियों से सूचना एकत्र की जाती है।

सर्वेक्षण के साधनों की तैयारी

सर्वेक्षणों में उपयोग किया जाने वाला सर्वाधिक प्रचलित साधन प्रश्नावली या साक्षात्कार अनुसूची है। प्रश्नावली या तो स्वयं उत्तरदाता द्वारा भरी जाती है या फिर शोधकर्ता (गणनाकार) अथवा प्रशिक्षित जाँचकर्ता द्वारा भरी जाती है। प्रश्नावली या साक्षात्कार अनुसूची तैयार करने में आपको निम्नलिखित बातों का ध्यान रखना चाहिए:

- *प्रश्नावली बहुत लम्बी नहीं होनी चाहिए।*
  जहाँ तक संभव हो सके, प्रश्नों की संख्या कम से कम होनी चाहिए। लंबी प्रश्नावली उत्तरदाताओं को हतोत्साहित करती है।
- *प्रश्नावली समझने में आसान होनी चाहिए।*
  अस्पष्ट या कठिन शब्दों से बचना चाहिए।
- *प्रश्न ऐसे क्रम में व्यवस्थित किए जाने चाहिए कि उत्तर देने वाला व्यक्ति आराम से उत्तर दे सके।*
- *प्रश्नावली सामान्य प्रश्नों से आरम्भ होकर विशिष्ट प्रश्नों की ओर बढ़नी चाहिए।*
  प्रश्नावली की शुरुआत सामान्य प्रश्नों के साथ होनी चाहिए और विशिष्ट प्रश्न क्रमशः बाद में दिए जाने चाहिए। इससे उत्तरदाता निश्चिन्त हो जाता है। उदाहरणार्थ:

गलत प्रश्न

(क) क्या बिजली के प्रभार में वृद्धि को उचित ठहराया जा सकता है?

(ख) क्या आपके क्षेत्र में बिजली की पूर्ति नियमित रहती है?

सही प्रश्न

(क) क्या आपके क्षेत्र में बिजली की पूर्ति नियमित रहती है?

(ख) क्या बिजली के प्रभार में वृद्धि को उचित ठहराया जा सकता है?

- *प्रश्न यथातथ्य एवं स्पष्ट होने चाहिए।* उदाहरणार्थ:

गलत प्रश्न

आप आकर्षक दिखने के लिए अपनी आय का कितना प्रतिशत भाग कपड़ों पर खर्च करते हैं?

सही प्रश्न

आप अपनी आय का कितना प्रतिशत भाग कपड़ों पर खर्च करते हैं?

- *प्रश्न अनेकार्थक या अस्पष्ट नहीं होने चाहिए। प्रश्न ऐसे हों ताकि उत्तरदाता शीघ्र, सही एवं स्पष्ट उत्तर देने में सक्षम रहे।* उदाहरणार्थ:

गलत प्रश्न

क्या आप प्रतिमाह पुस्तकों पर बहुत पैसा खर्च करते हैं?

सही प्रश्न

सही विकल्प पर सही (✓)का निशान लगाएँ।

आप प्रतिमाह पुस्तकों पर कितना खर्च करते हैं?

(क) 200/- रु से कम

(ख) 200/- से 300/- रु के बीच

(ग) 300/- से 400/- रु के बीच

(घ) 400/- रु से अधिक

- *प्रश्न दोहरी नकारात्मकता वाले नहीं होने चाहिए।* प्रश्नों को 'क्या आप नहीं' से शुरू नहीं करना चाहिए, क्योंकि इनसे पूर्वाग्रह-ग्रस्त उत्तर मिलने की संभावना हो सकती है। उदाहरणार्थ:

गलत प्रश्न

क्या आप ऐसा नहीं सोचते कि धूम्रपान को निषिद्ध किया जाना चाहिए।

सही प्रश्न

© NCERT not to be republished

क्या आप सोचते हैं कि धूम्रपान को निषिद्ध किया जाना चाहिए?

- *प्रश्न संकेतक प्रश्न नहीं होने चाहिए, जिससे उत्तरदाता को जवाब देने के लिए सूत्र मिले।* उदाहरणार्थ:

गलत प्रश्न

क्या आप इस उच्च कोटि की चाय के स्वाद को पसंद करते हैं?

सही प्रश्न

आपको इस चाय का स्वाद कैसा लगा?

- प्रश्न से उत्तर के विकल्प का संकेत नहीं मिलना चाहिए? उदाहरणार्थ:

गलत प्रश्न

क्या आप कॉलेज के बाद नौकरी करना चाहेंगी या गृहिणी बनना चाहेंगी?

सही प्रश्न

आप कॉलेज के बाद क्या करना चाहेंगी?

प्रश्नावली में *परिमितोत्तर* (संरचित) प्रश्न या *मुक्तोत्तर* (असंरचित) प्रश्न हो सकते हैं। उपरोक्त प्रश्न, कि एक विद्यार्थी कॉलेज के बाद क्या करना चाहता है, एक मुक्तोत्तर प्रश्न है।

संरचित प्रश्न या असंरचित प्रश्न या तो द्विविध प्रश्न हो सकते हैं या फिर बहुविकल्पी प्रश्न हो सकते हैं। जब किसी प्रश्न के उत्तर में 'हाँ' या 'नहीं' के मात्र दो ही विकल्प होते हैं तो इसे द्विविध प्रश्न कहते हैं।

जब प्रश्नावली के अंतर्गत दो से अधिक उत्तरों के विकल्प होते हैं, वहाँ *बहुविकल्पी* प्रश्न अधिक उपयुक्त होते हैं। उदाहरणार्थ,

प्रश्न – आपने अपनी ज़मीन क्यों बेंच दी?

(क) कर्ज चुकाने के लिए।

(ख) बच्चों की शिक्षा हेतु धन की व्यवस्था के लिए।

(ग) किसी अन्य संपत्ति में निवेश हेतु।

(घ) कोई अन्य कारण (कृपया स्पष्ट करें)।

मुक्तोत्तर प्रश्न विश्लेषण की दृष्टि से उपयोग, स्कोर तथा कोड के लिए आसान होते हैं, क्योंकि उत्तरदाताओं को दिए गए विकल्पों में से उत्तर चुनना होता है। लेकिन इनके उपयुक्त विकल्प लिखने में कठिनाई होती है। इन विकल्पों को स्पष्ट तौर से लिखा जाना चाहिए ताकि मुद्दे के दोनों पहलुओं का प्रतिनिधित्व हो सके। यहाँ पर एक संभावना यह भी रहती है कि व्यक्ति-विशेष का सही उत्तर, दिए गए विकल्पों में से कोई भी न हो। इस के लिए *कोई अन्य* का विकल्प दिया जाना चाहिए, जहाँ उत्तरदाता अपना वह उत्तर लिख सके, जिसकी अपेक्षा शोधकर्ता/सर्वेक्षक को भी नहीं होती। इसके अलावा, बहु-विकल्पी प्रश्नों की एक अन्य सीमा यह भी है कि इसमें उत्तरदाता को अनेक वैकल्पिक उत्तर देकर प्रतिबंधित कर दिया जाता है, अन्यथा उत्तरदाता इन विकल्पों से भिन्न उत्तर भी दे सकता था।

मुक्तोत्तर प्रश्न के अंतर्गत व्यक्ति को उत्तर देने की अधिक व्यक्तिगत छूट रहती है, लेकिन इनकी सही व्याख्या करने में कठिनाई होती है तथा इन्हें स्कोर करने में मुश्किल होती है, चूँकि उत्तरों में काफी विभिन्नता होती है। उदाहरणार्थ,

प्रश्न – वैश्वीकरण के विषय में आपके क्या विचार हो सकते है?

आँकड़ा-संग्रह की विधि

क्या आपने कोई ऐसा टेलीविजन शो देखा है, जिसमें रिपोर्टर ने बच्चों, गृहणियों या आम जनता से क्रमशः उनकी परीक्षा या साबुन के किसी ब्रांड या किसी राजनीतिक पार्टी के बारे में प्रश्न पूछा हो? इन प्रश्नों के पूछने का उद्देश्य आँकड़ा-संग्रह करने के लिए सर्वेक्षण करना है। आँकड़ा-संग्रह की तीन आधारभूत विधियाँ हैं: (क) वैयक्तिक साक्षात्कार (ख) डाक द्वारा सर्वेक्षण (प्रश्नावली भेजना) और (ग) टेलीफोन-साक्षात्कार।

© NCERT not to be republished

वैयक्तिक साक्षात्कार

यह विधि तभी उपयोग में लाई जाती है जब शोधकर्ता सभी सदस्यों के पास जा सकता हो। इसमें शोधकर्ता (जाँचकर्ता) आमने-सामने होकर उत्तरदाता से साक्षात्कार करता है।

वैयक्तिक साक्षात्कारों को कई कारणों से प्राथमिकता दी जाती है। इसमें सर्वेक्षक एवं उत्तरदाता के बीच व्यक्तिगत संपर्क होता है। सर्वेक्षक या साक्षात्कारकर्त्ता को यह अवसर मिलता है कि वह उत्तरदाता को अध्ययन के उद्देश्य के बारे में बता सके तथा उत्तरदाता की किसी भी पूछताछ का जवाब दे सके। इसमें साक्षात्कारकर्त्ता उत्तरदाता से यह निवेदन कर सकता है कि वह विशेष महत्व के बिंदुओं को विस्तार से बताए। इससे अपनिर्वचन (गलत व्याख्या) तथा गलतफहमी से बचा जा सकता है। साथ ही उत्तरदाता की प्रतिक्रियाओं को देख कर कुछ संपूरक सूचनाएँ भी प्राप्त हो सकती हैं।

वैयक्तिक साक्षात्कार की कुछ कमियाँ भी हैं। यह काफी खर्चीली होती है तथा इसमें प्रशिक्षित साक्षात्कार कर्त्ताओं की ज़रूरत होती है। इसमें सर्वेक्षण पूरा करने में काफी अधिक समय लगता है। कभी-कभी शोधकर्ता/सर्वेक्षक की उपस्थिति के कारण उत्तरदाता सही बात नहीं भी बताते हैं।

डाक द्वारा प्रश्नावली भेजना

जब सर्वेक्षण में आँकड़ों को डाक द्वारा संगृहीत किया जाता है, तो प्रत्येक उत्तरदाता को डाक द्वारा प्रश्नावली इस निवेदन के साथ भेजी जाती है कि वह इसे पूरी कर एक निश्चित तारीख तक वापस अवश्य भेज दे। इस का सबसे बड़ा लाभ है कि यह बहुत कम खर्चीली होती है। इसके साथ ही इस विधि के द्वारा शोधकर्त्ता/सर्वेक्षक काफी दूर-दराज के क्षेत्रों तक पहुँच सकते हैं, जो संभवत: व्यक्ति या टेलीफोन की पहुँच से भी बाहर हो सकते हैं। इस विधि में साक्षात्कारकर्त्ता उत्तरदाताओं पर प्रभाव भी नहीं डाल पाते। साथ ही, यह उत्तरदाता को पर्याप्त समय देता है ताकि, वह सोच-विचार कर प्रश्नों के उत्तर दे सके। आजकल *आन लाइन* सर्वेक्षण या **संक्षिप्त संदेश सेवा** (SMS) द्वारा सर्वेक्षण काफी लोकप्रिय हो रहे हैं। क्या आप जानते हैं कि ऑन लाइन सर्वेक्षण कैसे आयोजित किए जाते हैं?

डाक द्वारा सर्वेक्षण की यह कमी है कि प्रश्नावली के निर्देशों के स्पष्टीकरण के अवसर नहीं मिलते हैं। अत: इसमें प्रश्न की अपनिर्वचन की संभावना रहती है। साथ ही डाक-सर्वेक्षण द्वारा कम संख्या में उत्तरदाताओं से उत्तर प्राप्ति की भी संभावना रहती है, क्योंकि प्रश्नावली को बिना पूरा भरे ही लौटाने की या प्रश्नावली को बिल्कुल ही न लौटाने की भी संभावना रहती है और साथ ही डाक विभाग द्वारा प्रश्नावली के खो जाने आदि की संभावना भी रहती है।

टेलीफोन साक्षात्कार

टेलीफोन साक्षात्कार के अंतर्गत शोधकर्त्ता/जाँचकर्त्ता टेलीफोन के माध्यम से सर्वेक्षण करता है। टेलीफोन साक्षात्कार का लाभ है कि यह वैयक्तिक साक्षात्कार की अपेक्षा सस्ता होता है और इसे कम समय में ही सम्पन्न किया जा सकता है। यह प्रश्नों को स्पष्ट कर सर्वेक्षक/ शोध कर्ता के लिए उत्तरदाता की मदद करने में सहायक होता है। टेलीफोन साक्षात्कार उन मामलों में अधिक बेहतर होता है, जहाँ वैयक्तिक साक्षात्कार के समय उत्तरदाता कुछ खास प्रश्नों के उत्तर देने में झिझक महसूस करता है।

© NCERT not to be republished

इस विधि की कमी यह है कि इसमें लोगों तक सर्वेक्षक की पहुँच सीमित हो जाती है, क्योंकि बहुत से लोगों के पास निजी टेलीफोन नहीं भी हो सकते हैं। इसके साथ टेलीफोन साक्षात्कार की कमी यह भी है, कि संवेदनशील मुद्दों पर उत्तरदाताओं की उन प्रतिक्रियाओं को दृश्य रूप में नहीं देखा जा सकता है, जो इन विषयों पर सही जानकारी प्राप्त करने में सहायक होती हैं।

क्रियात्मक गतिविधियाँ

- आपको एक ऐसे व्यक्ति से जानकारी (सूचनाएँ) प्राप्त करनी है जो भारत के दूर-दराज के गाँव में रहता है। इस व्यक्ति से सूचना प्राप्त करने के लिए आँकड़ा-संग्रह की कौन सी विधि सर्वाधिक उपयुक्त रहेगी और क्यों? विवेचना कीजिए।
- आपको किसी विद्यालय की अध्ययन गुणवत्ता के बारे में अध्यापक से एक साक्षात्कार करना है। यदि वहाँ पर विद्यालय का प्रधानाचार्य उपस्थित है, तो किस प्रकार की समस्याएँ पैदा हो सकती हैं?

प्रायोगिक सर्वेक्षण

एक बार जब सर्वेक्षण हेतु प्रश्नावली तैयार हो जाए तो यह सलाह दी जाती है कि एक छोटे समूह का सर्वेक्षण करके देख लिया जाना चाहिए, जिसे *प्रायोगिक सर्वेक्षण* के रूप में या प्रश्नावली की *पूर्व-परीक्षा* के रूप में जाना जाता है। सर्वेक्षण के बारे में प्रारंभिक अनुमान लगाने में प्रायोगिक सर्वेक्षण सहायक होता है। यह प्रश्नावली के पूर्व-परीक्षण में भी सहायक होता है, ताकि प्रश्नों की कमियों एवं त्रुटियों को पता किया जा सके। इसके साथ ही प्रायोगिक सर्वेक्षण प्रश्नों की उपयुक्तता, निर्देशों की स्पष्टता, सर्वेक्षक (गणनाकार) की कार्य-दक्षता तथा वास्तविक सर्वेक्षण में आनेवाली लागत एवं समय का अनुमान लगाने में भी सहायता करता है।

4. जनगणना तथा प्रतिदर्श सर्वेक्षण

जनगणना या पूर्ण गणना (Census or Complete Enumeration)

वह सर्वेक्षण, जिसके अंतर्गत जनसंख्या के सभी तत्व शमिल होते हैं, उसे *जनगणना या पूर्ण गणना की विधि*

लाभ	*हानि*
• उच्चतम उत्तर दर • सभी प्रकार के प्रश्नों के उपयोग की छूट • मुक्तोत्तर प्रश्नों के उपयोग के लिए बेहतर • अस्पष्ट प्रश्नों के लिए स्पष्टीकरण का अवसर	• बहुत खर्चीला • उत्तरदाता को प्रभावित करने की संभावना • अधिक समय लेने वाला
• कम खर्चीला • ग्रामीण तथा सुदूर क्षेत्रों तक पहुँच का एकमात्र साधन • उत्तरदाता पर कोई प्रभाव नहीं • उत्तरदाता की गोपनीयता सुरक्षित • संवेदनशील मुद्दों के लिए सर्वोत्तम	• निरक्षरों के द्वारा उपयोग संभव नहीं • उत्तर प्राप्त करने में अधिक समय • अस्पष्ट प्रश्नों के स्पष्टीकरण का अभाव • प्रतिक्रियाएँ देखना सम्भव नहीं
• अपेक्षाकृत कम खर्चीला • उत्तरदाता को प्रभावित करने की अपेक्षाकृत कम संभावना • अपेक्षाकृत उच्च उत्तर दर।	• सीमित उपयोग • प्रतिक्रियाएँ देखना सम्भव नहीं • उत्तरदाताओं को प्रभावित करने की संभावनाएँ

© NCERT not to be republished

कहा जाता है। यदि कुछ खास संस्थाएँ भारत की संपूर्ण जनसंख्या के बारे में अध्ययन की रुचि रखती हैं, तो उन्हें भारत के सभी शहरों एवं गाँवों के सभी परिवारों के बारे में जानकारी प्राप्त करनी होगी। इस विधि की प्रमुख विशेषता है कि इसके अंतर्गत संपूर्ण जनसंख्या की प्रत्येक व्यष्टिगत इकाई को सम्मिलित करना होता है। आप ऐसा नहीं कर सकते हैं कि कुछ इकाइयों को चुन लें और कुछ को छोड़ दें। आपने संभवत: *भारत की जनगणना* के बारे में सुना होगा, जो हर दस साल में एक बार होती है। इसके अंतर्गत घर-घर जाकर जानकारी ली जाती है और पूरे भारत के हर एक परिवार को इसमें सम्मिलित किया जाता है। इसके अंतर्गत जन्म एवं मृत्युदर, साक्षरता, रोजगार, आयु संभाविता या प्रत्याशित आयु, जनसंख्या के आकार एवं सरंचना आदि के जनसांख्यिकीय आँकड़े जुटाए जाते हैं, जिन्हें भारत के महानिदेशक द्वारा संगृहीत एवं प्रकाशित किया जाता है। *भारत में पिछली जनगणना* 2011 में की गई थी।

जनगणना 2001 के अनुसार भारत की जनसंख्या 2011 की जनगणना के अनुसार भारत की जनसंख्या 121.09 करोड़ है, जो 2001 में 102.87 करोड़ थी। 1901 की जनगणना ने देश की जनसंख्या 23.83 करोड़ दर्शाई थी। तब से 110 वर्षों की समयावधि में, देश की जनसंख्या 97 करोड़ से भी अधिक बढ़ गई है। जनसंख्या की औसत वार्षिक वृद्धि दर, जो 1971-81 में 2.2 प्रतिशत प्रतिवर्ष थी, 1991-2001 में घटकर 1.97 प्रतिशत हो गई तथा 2001-2011 में 1.64 हो गई।

जनसंख्या तथा प्रतिदर्श (Population and Sample)

सांख्यिकी में *'समष्टि'* शब्द का तात्पर्य है अध्ययन-क्षेत्र के अंतर्गत आने वाली सभी मदों/इकाइयों की समग्रता। अत: *समष्टि* एक ऐसा समूह है, जिस पर किसी अध्ययन के परिणाम लागू हो सकें। सर्वेक्षण के उद्देश्य के अनुसार किसी *समष्टि के अंतर्गत सदैव ऐसी सभी व्यष्टि तथा इकाइयाँ/मदें आती हैं, जिनमें कुछ विशेषताएँ (या विशेषताओं का समूह)* हों। प्रतिदर्श चुनने में पहला कार्य *समष्टि* की पहचान करना है। एक बार जब *समष्टि* की पहचान हो जाती है तो शोधकर्ता इसका अध्ययन करने का एक तरीका चुनता है। यदि शोधकर्ता को लगता है कि समूची समष्टि या जनसंख्या का सर्वेक्षण *संभव नहीं है, तो वह एक प्रतिनिधि प्रतिदर्श चुन सकता है। एक आदर्श प्रतिदर्श (प्रतिनिधि प्रतिदर्श)* सामान्यत: समष्टि से छोटा होता है तथा अपेक्षाकृत कम लागत एवं कम समय में समष्टि के बारे में पर्याप्त सही सूचनाएँ प्रदान करने में सक्षम होता है।

मान लें कि आप, किसी क्षेत्र-विशेष के लोगों की औसत आय के बारे में अध्ययन करना चाहते हैं। गणनाविधि के अनुसार, आपको उस क्षेत्र के प्रत्येक व्यक्ति की आय का पता करने के बाद उनका कुल

© NCERT not to be republished

योग करके वहाँ के लोगों की संख्या से भाग देकर वहाँ के लोगों की औसत आय पता करनी होगी। इस विधि के अंतर्गत बहुत खर्च आएगा, क्योंकि इसके लिए भारी संख्या में परिगणकों की भर्ती करनी होती है। इसके विकल्प के रूप में, आप उस क्षेत्र के, कुछ व्यक्तियों का *प्रतिदर्श चुन कर* उनकी आय को जान लेते हैं। चुने गए समूह के व्यक्तियों की औसत आय ही उस पूरे क्षेत्र के लोगों की औसत आय होती है। उदाहरण के लिए–

- *शोध समस्या:* मणिपुर राज्य के चूराचाँदपुर जिले के कृषि श्रमिकों की आर्थिक स्थिति का अध्ययन करना।
- *समष्टि:* चूराचाँदपुर जिले के समस्त कृषि श्रमिक।
- *प्रतिदर्श (नमूना):* चूराचाँदपुर जिले के 10 प्रतिशत कृषि श्रमिक।

अधिकतर सर्वेक्षण प्रतिदर्श सर्वेक्षण ही होते हैं। सांख्यिकी में इन्हें कई कारणों से प्राथमिकता दी जाती है। यह प्रतिदर्श कम खर्च में एवं अल्प समय में पर्याप्त विश्वसनीय एवं सही सूचनाएँ उपलब्ध करा सकते हैं। प्रतिदर्श, चूँकि समष्टि से छोटा होता है, अत: सघन पूछताछ के द्वारा अधिक विस्तृत सूचनाएँ संगृहीत की जा सकती हैं। इसके लिए परिगणकों की छोटी टोली की ही जरूरत होगी, जिन्हें आसानी से प्रशिक्षित किया जा सकता है तथा उनके कार्य की निगरानी भली-भाँति की जा सकती है। अब प्रश्न यह उठता है कि इस प्रतिदर्श का चयन कैसे करें? प्रतिदर्श चयन के दो प्रचलित तरीके हैं, जिन्हें यादृच्छिक एवं अयादृच्छिक प्रतिदर्श कहते हैं। इन दोनों प्रकार के प्रतिदर्शों के अंतर का विवरण आगे दिया जा रहा है।

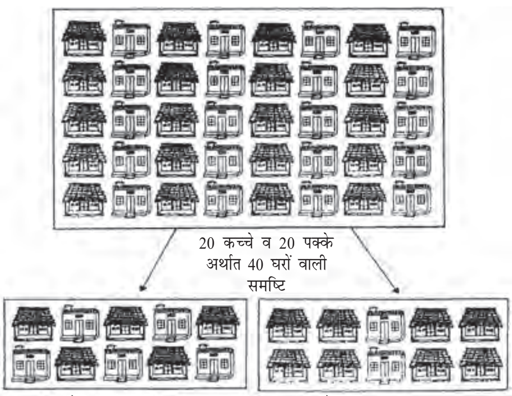
क्रियात्मक गतिविधियाँ

- भारत एवं चीन में अगली जनगणना किन-किन वर्षों में की जाएगी?
- यदि आप XI वीं कक्षा की अर्थशास्त्र की नई पाठ्यपुस्तक के बारे में छात्रों की राय जानना चाहते हैं, तो आप की समष्टि क्या होगी और प्रतिदर्श क्या होगा?
- यदि कोई शोधकर्ता पंजाब में गेहूँ की फसल के औसत उत्पादन का आकलन करना चाहता है, तो उसकी समष्टि और प्रतिदर्श समूह क्या होंगे?

यादृच्छिक प्रतिचयन

जैसा कि नाम से स्पष्ट है, यादृच्छिक प्रतिचयन वह होता है, जहाँ समष्टि प्रतिदर्श-समूह से व्यष्टिगत इकाइयों (प्रतिदर्श) को यादृच्छिक रूप से चुना जाता है। मान लें कि सरकार एक क्षेत्र-विशेष में रहने वाले परिवारों के बजट पर पेट्रोलियम पदार्थों की कीमतों की वृद्धि के प्रभाव की जाँच करना चाहती है। इसके लिए 30 परिवारों का प्रतिनिधिक (यादृच्छिक) प्रतिदर्श प्राप्त करके उसका अध्ययन करना है। इनके चुनाव के लिए पहले उस क्षेत्र के सभी 300 परिवारों के नाम पर्चियों पर लिखे जाते हैं और फिर उन पर्चियों को पूरी तरह आपस में मिला दिया जाता है। इसके बाद, उसमें से साक्षात्कार के लिए बारी-बारी से 30 नाम पर्चियों द्वारा चुन लिए जाते हैं।

यादृच्छिक प्रतिचयन में प्रत्येक व्यक्ति के चुने जाने की समान संभावना होती है और चुना गया



5 पक्के और 5 कच्चे घरों का प्रतिनिधि प्रतिदर्श 2 पक्के और 8 कच्चे घरों का अप्रतिनिधि प्रतिदर्श

© NCERT not to be republished

व्यक्ति ठीक वैसा ही होता है, जैसा कि नहीं चुना गया व्यक्ति। उपर्युक्त उदाहरण में 300 प्रतिदर्श इकाइयों *(प्रतिचयन रचना)* की समष्टि में सभी इकाइयों को, 30 इकाइयों के प्रतिदर्श में, चुने जाने का समान अवसर प्राप्त हुआ। अतः इस तरह से निकाले गए प्रतिदर्श को ही यादृच्छिक प्रतिदर्श कहा जाता है। इस विधि को लाटरी विधि के नाम से भी जाना जाता है। आजकल यादृच्छिक प्रतिदर्शों के चयन के लिए कंप्यूटर प्रोग्राम का उपयोग किया जाता है।

निर्गम निर्वाचन (Exit Poll)

आपने देखा ही होगा कि जब चुनाव होते हैं तो टेलीविजन नेटवर्क चुनाव संबंधी समाचार दिखाते हैं। इसके साथ ही ये लोग इसका पूर्वानुमान भी दिखाते हैं कि कौन सी पार्टी जीत सकती है। इसे निर्गम मतदान (ऐग्जिट पोल) के रूप में किया जाता है। इसके अंतर्गत मतदान केंद्रों से मतदान करके निकलने वाले मतदाताओं से यादृच्छिक प्रतिदर्श लेने के लिए पूछा जाता है कि उन्होंने किसे मत दिया है। यहाँ मतदाताओं के प्रतिदर्श द्वारा प्राप्त आँकड़ों से चुनाव जीतने वालों के बारे में पूर्वानुमान लगाया जाता है। आपने देखा होगा कि निर्गम निर्वाचन सदैव सही अनुमान नहीं लगाते हैं। क्यों?

क्रियात्मक गतिविधि

- आपको भारत में खाद्यान्न उत्पादन की पिछले 50 वर्षों की प्रवृत्ति का विश्लेषण करना है। चूँकि सभी 50 वर्षों के लिए आँकड़े एकत्रित करना मुश्किल है, अतः आपको 10 वर्षों के एक प्रतिदर्श का चयन करना है। आप यादृच्छिक संख्या सारणी का प्रयोग करते हुए अपने प्रतिदर्श कैसे चुनेंगे?

अयादृच्छिक प्रतिचयन

ऐसी स्थिति भी हो सकती है जब आपको किसी क्षेत्र के 100 परिवारों में से 10 को चुनना हो। आपको यह तय करना है कि किन परिवारों को चुनें और किन्हें छोड़ दें। आप ऐसे घरों को चुन सकते हैं, जो आपके लिए सुविधा जनक हों या फिर अपने मित्र या परिचित के घर को चुन सकते हैं। इस मामले में, आप 10 परिवारों को चुनने के लिए अपने निर्णय (पूर्वाग्रह) का प्रयोग करते हैं। ऐसी स्थिति में 100 परिवारों में से आपके द्वारा चुने गए 10 परिवार, अयादृच्छिक प्रतिदर्श द्वारा नहीं चुने गए हैं। *अतः किसी अयादृच्छिक प्रतिदर्श में उस समष्टि की सभी इकाइयों के चुने जाने की समान संभावनाएँ नहीं होती हैं और इसमें सर्वेक्षक की सुविधा या निर्णय की भूमिका महत्वपूर्ण हो जाती है।* इन्हें चूँकि प्रायः अपने निर्णय, उद्देश्य, सुविधा तथा नियतमात्रा (कोटा) के आधार पर चुना जाता है, अतः इसे अयादृच्छिक प्रतिदर्श के रूप में जाना जाता है।

5. प्रतिचयन एवं अप्रतिचयन त्रुटियाँ

प्रतिचयन त्रुटियाँ (Sampling Errors)

संख्यात्मक मानों वाली जनसंख्या की दो महत्वपूर्ण विशेषताएँ होती हैं जो यहाँ सुसंगत हैं। पहली, केंद्रीय प्रवृत्ति, जिसका मापन औसत (मध्यमान), माध्य या बहुलक के द्वारा किया जा सकता है। दूसरी, विचलन, जिसका मापन 'मानक विचलन', 'माध्य विचलन', परास आदि की गणना द्वारा किया जा सकता है।

प्रतिदर्श का उद्देश्य जनसंख्या प्राचलों के एक या अधिक आकलनों को प्राप्त करना होता है। प्रतिचयन त्रुटि प्रतिदर्श आकलन तथा उसी के जनसंख्या प्राचल (समष्टि विशेष जैसे औसत आय आदि के वास्तविक मूल्य) में अंतर को इंगित करता है।

उदाहरणार्थ–

मणिपुर के 5 कृषकों की आमदनी का उदाहरण लें। मान लें, चर **X** (कृषकों की आमदनी) के मापन 500, 550, 600, 650, 700 हैं। हमने देखा कि यहाँ समष्टि का औसत (500 + 550 + 600 + 650 + 700) ÷ 5 = 3000 ÷ 5 = 600 है।

© NCERT not to be republished

अब मान लीजिए हम दो व्यक्तियों का एक ऐसा प्रतिदर्श चुनते हैं जहाँ **X** के मूल्य 500 एवं 600 हैं। अब प्रतिदर्श का औसत (500+600)÷ 2 = 1100 ÷ 2 = 550 होता है।

यहाँ *आकलन की प्रतिचयन त्रुटि है*= 600 (असली मान) –550 (आकलन) = 50

अप्रतिचयन त्रुटियाँ (Non sampling Errors)

अप्रतिचयन त्रुटियाँ प्रतिचयन त्रुटियों की अपेक्षा अधिक गंभीर होती हैं। ऐसा इसलिए होता है कि प्रतिचयन त्रुटियों को बड़े आकार के प्रतिदर्श लेकर कम किया जा सकता है, पर अयादृच्छिक त्रुटियों को कम करना असंभव है, चाहे प्रतिदर्श का आकार बड़ा ही क्यों न रखा जाए। यहाँ तक कि जनगणना में भी अयादृच्छिक त्रुटि की संभावना हो सकती है। अयादृच्छिक त्रुटियों के कुछ उदाहरण निम्नलिखित हैं:

आँकड़ा अर्जन में त्रुटियाँ

इस प्रकार की त्रुटियाँ गलत उत्तरों को रेकार्ड करने से पैदा होती है। मान लीजिए, एक शिक्षक कक्षा के छात्रों से अध्यापक की मेज की लंबाई को मापने के लिए कहता है। छात्रों द्वारा लिए गए माप में अंतर हो सकते हैं। ये अंतर फीते में अंतर, छात्रों की लापरवाही, आदि के कारण हो सकते हैं। इसी प्रकार, मान लें कि हम संतरों की कीमत के बारे में आँकड़े एकत्र करना चाहते हैं। हम जानते हैं कि अलग-अलग दुकानों में तथा अलग-अलग बाजारों में संतरों की कीमत भिन्न-भिन्न हो सकती है। इसके साथ ही गुणवत्ता के आधार पर भी मूल्यों में अंतर हो सकता है। इसीलिए हम यहाँ पर केवल औसत कीमत को ही लेते हैं। रिकार्ड करने में त्रुटियों की संभावना रहती है, जब सर्वेक्षक या उत्तरदाता गलत आँकड़े रेकार्ड करता है या लिखता है। उदाहरण के लिए 31 को गलती से 13 लिखा जा सकता है।

अनुत्तर संबंधी त्रुटियाँ

अनुत्तर संबंधी त्रुटियों की संभावना तब होती है, जब साक्षात्कारकर्त्ता प्रतिदर्श सूची में सूचीबद्ध उत्तरदाता से संपर्क नहीं स्थापित कर पाता है या प्रतिदर्श सूची का कोई व्यक्ति उत्तर देने से मना कर देता है। ऐसे मामलों में प्रतिदर्श प्रेक्षण को प्रतिनिधि प्रतिदर्श नहीं माना जा सकता है।

प्रतिदर्श अभिनति (Sampling Bias)

प्रतिदर्श अभिनति (पूर्वाग्रह) की संभावना तब होती है जब प्रतिचयन योजना ऐसी हो कि उसके अंतर्गत समष्टि से कुछ ऐसे सदस्यों के सम्मिलित होने की संभावना नहीं है, जिन्हें प्रतिदर्श में शामिल किया जाना चाहिए था।

6. भारत की जनगणना तथा राष्ट्रीय प्रतिदर्श सर्वेक्षण (NSS)

राष्ट्रीय एवं राज्य दोनों ही स्तरों पर ऐसी संस्थाएँ होती हैं, जो सांख्यिकीय आँकड़ों को संगृहीत, संसाधित तथा सारणीकृत करती हैं। इनमें से राष्ट्रीय स्तर की कुछ प्रमुख संस्थाएँ हैं, सेन्सस ऑफ इंडिया, राष्ट्रीय प्रतिदर्श सर्वेक्षण (NSS), केंद्रीय सांख्यिकीय कार्यालय (CSO), भारत का महापंजीकार (RGI), वाणिज्यिक सतर्कता एवं सांख्यिकी महानिदेशालय (DGCIS) तथा श्रम ब्यूरो आदि।

केन्द्रीय सांख्यिकी संगठन जनसंख्या संबंधित सर्वाधिक पूर्ण एवं सतत जनसांख्यिकीय अभिलेख उपलब्ध कराती है। वर्ष 1881 के बाद से प्रत्येक 10 वर्ष के अंतराल पर नियमित जनगणना की जाती है। देश की आजादी के बाद पहली जनगणना वर्ष 1951 में हुई थी। इन जनगणनाओं के अंतर्गत जनसंख्या के विभिन्न पहलुओं के बारे में सूचनाएँ एकत्र की जाती हैं, जैसे आकार, घनत्व, लिंग-अनुपात, साक्षरता, स्थानांतरण तथा जनसंख्या का ग्रामीण-शहरी वितरण आदि।

© NCERT not to be republished

जनगणना आँकड़ों का निर्वचन एवं विश्लेषण भारत में अनेक आर्थिक और सामाजिक मुद्दों को समझने के लिए किया जाता है। राष्ट्रीय प्रतिदर्श सर्वेक्षण संगठन की स्थापना भारत सरकार द्वारा समाज-आर्थिक मुद्दों पर राष्ट्रीय स्तर के सर्वेक्षणों के लिए की गई थी। यह संगठन बारी-बारी से निरंतर सर्वेक्षण करता रहता है। इस संगठन के सर्वेक्षणों द्वारा संग्रह किए गए आँकड़े समय-समय पर विभिन्न रिपोर्टों एवं इसकी त्रैमासिक पत्रिका *'सर्वेक्षण'* में प्रकाशित किए जाते हैं। ये आँकड़े मूलतः सामाजिक-आर्थिक मुद्दों पर होते हैं। इसके साथ ही राष्ट्रीय प्रतिदर्श सर्वेक्षण संगठन साक्षरता, विद्यालयी नामांकन, शैक्षिक सेवाओं का समुपयोजन, रोजगार, बेरोजगारी, विनिर्माण एवं सेवा क्षेत्रकों के उद्यमों, रुग्णता, मातृत्व, शिशु-देखभाल और सार्वजनिक वितरण प्रणाली के समुपयोजन आदि पर भी अनुमानित आँकड़े उपलब्ध कराता है। राष्ट्रीय प्रतिदर्श सर्वेक्षण (NSS) का 60वाँ क्रमिक सर्वेक्षण (जनवरी-जून, 2004) अस्वस्थता तथा स्वास्थय सेवाओं पर था। राष्ट्रीय प्रतिदर्श सर्वेक्षण (NSS) का 68वाँ क्रमिक सर्वेक्षण (2011-12) उपभोक्ता व्यय पर था। साथ ही राष्ट्रीय प्रतिदर्श सर्वेक्षण संगठन उद्योगों का वार्षिक सर्वेक्षण, फसल अनुमान सर्वेक्षण आदि का भी आयोजन करता है। यह उपभोक्ता कीमत सूचकांक से संबंधित संख्याओं के संकलन के लिए ग्रामीण एवं शहरी खुदरा कीमतों का संग्रह आदि भी करता है।

7. सारांश

संख्याओं के रूप में व्यक्त किए गए आर्थिक तथ्य आँकड़े कहलाते हैं। आँकड़ों के संग्रह का उद्देश्य किसी समस्या और उसके कारणों को समझ कर उसकी व्याख्या एवं विश्लेषण करना है। प्राथमिक आँकड़ों का संग्रह सर्वेक्षण आयोजित करके किया जाता है। सर्वेक्षणों के कई चरण होते हैं, जिन्हें सावधानी पूर्वक नियोजित करने की आवश्यकता होती है। ऐसी अनेक संस्थाएँ हैं, जो इन सांख्यिकीय आँकड़ों का संग्रह, संसाधन, सारणीयन, तथा प्रकाशन करती हैं। इनका प्रयोग द्वितीयक आँकड़ों के रूप में किया जा सकता है। आँकड़ों के स्रोत का चुनाव एवं इनके संग्रह की विधा अध्ययन के उद्देश्य पर निर्भर करती है।

© NCERT not to be republished

पुनरावर्तन

- आँकड़े ऐसे साधन हैं, जो सूचनाएँ उपलब्ध कराकर किसी भी समस्या के विषय में ठोस निष्कर्ष पर पहुँचने में सहायता देती हैं।
- प्राथमिक आँकड़े व्यक्ति द्वारा स्वयं एकत्र की गई सूचनाओं पर निर्भर होते हैं।
- सर्वेक्षण वैयक्तिक साक्षात्कारों, डाक द्वारा प्रश्नावलियाँ भेजकर तथा टेलीफोन साक्षात्कार द्वारा किये जा सकते हैं।
- जनगणना के अंतर्गत समष्टि की सभी इकाइयों/व्यष्टियों को सम्मिलित किया जाता है।
- प्रतिदर्श, समष्टि से चयनित किया गया एक छोटा समूह होता है, जिसके द्वारा संबंधित सूचनाएँ प्राप्त की जा सकती हैं।
- यादृच्छिक प्रतिचयन के अंतर्गत प्रत्येक व्यक्ति को सूचना प्रदान करने हेतु चुने जाने के लिए समान अवसर दिया जाता है।
- प्रतिदर्श त्रुटियाँ वास्तविक समष्टि तथा इनके आकलन के बीच अंतर के कारण पैदा होती हैं।
- अप्रतिचयन त्रुटियाँ आँकड़ों के अर्जन के दौरान पैदा हो सकती हैं, जो उत्तर न देने के कारण, या चयन में पूर्वाग्रह के कारण हो सकती हैं।
- राष्ट्रीय स्तर पर 'भारत की जनगणना' तथा 'राष्ट्रीय प्रतिदर्श सर्वेक्षण संगठन' दो महत्वपूर्ण संस्थाएँ हैं, जो विभिन्न महत्वपूर्ण आर्थिक एवं सामाजिक मुद्दों पर आँकड़ों का संग्रहण, संसाधन तथा सारणीयन करती हैं।

अभ्यास

1. निम्नलिखित प्रश्नों के लिए कम से कम चार उपयुक्त बहु विकल्पी वाक्यों की रचना करें:
   (क) जब आप एक नई पोशाक खरीदें तो इनमें से किसे सबसे महत्वपूर्ण मानते है?
   (ख) आप कम्प्यूटर का इस्तेमाल कितनी बार करते हैं?
   (ग) निम्नलिखित में से आप किस समाचार पत्र को नियमित रूप से पढ़ते हैं?
   (घ) पेट्रोल की कीमत में वृद्धि न्यायोचित है?
   (ङ) आपके परिवार की मासिक आमदनी कितनी है?
2. पाँच द्विमार्गी प्रश्नों की रचना करें (हाँ / नहीं के साथ)।
3. सही विकल्प को चिह्नित करें:
   (क) आँकड़ों के अनेक स्रोत होते हैं (सही / गलत)।
   (ख) आँकड़ा-संग्रह के लिए टेलीफोन सर्वेक्षण सर्वाधिक उपयुक्त विधि है, विशेष रूप से जहाँ पर जनता निरक्षर हो और दूर-दराज के काफी बड़े क्षेत्रों में फैली हो (सही / गलत)।
   (ग) सर्वेक्षक/शोधकर्ता द्वारा संग्रह किए गए आँकड़े द्वितीयक आँकड़े कहलाते हैं (सही / गलत)।
   (घ) प्रतिदर्श के अयादृच्छिक चयन में पूर्वाग्रह (अभिनति) की संभावना रहती है (सही / गलत)।
   (ङ) अप्रतिचयन त्रुटियों को बड़ा प्रतिदर्श अपनाकर कम किया जा सकता है (सही / गलत)।

© NCERT not to be republished

4. निम्नलिखित प्रश्नों के बारे में आप क्या सोचते हैं? क्या आपको इन प्रश्नों में कोई समस्या दीख रही है? यदि हाँ, तो कैसे?

   (क) आप अपने सबसे नजदीक के बाजार से कितनी दूर रहते है?
   (ख) यदि हमारे कूड़े में प्लास्टिक थैलियों की मात्रा 5 प्रतिशत है तो क्या इन्हें निषेधित किया जाना चाहिए?
   (ग) क्या आप पेट्रोल की कीमत में वृद्धि का विरोध नहीं करेंगे?
   (घ) क्या आप रासायनिक उर्वरक के उपयोग के पक्ष में हैं?
   (ङ) क्या आप अपने खेतों में उर्वरक इस्तेमाल करते हैं?
   (च) आपके खेत में प्रति हेक्टेयर कितनी उपज होती है?

5. आप बच्चों के बीच शाकाहारी आटा नूडल की लोकप्रियता का अनुसंधान करना चाहते हैं। इस उद्देश्य से सूचना-संग्रह करने के लिए एक उपयुक्त प्रश्नावली बनाएँ?

6. 200 फार्म वाले एक गाँव में फसल उत्पादन के स्वरूप पर एक अध्ययन आयोजित किया गया। इनमें से 50 फार्मों का सर्वेक्षण किया गया, जिनमें से 50 प्रतिशत पर केवल गेहूँ उगाए जाते हैं। समष्टि एवं प्रतिदर्श के आकार क्या हैं?

7. प्रतिदर्श, समष्टि तथा चर के दो-दो उदाहरण दें।

8. इनमें से कौन सी विधि द्वारा बेहतर परिणाम प्राप्त होते हैं, और क्यों?
   (क) गणना (जनगणना) (ख) प्रतिदर्श

9. इनमें कौन सी त्रुटि अधिक गंभीर है और क्यों?

   (क) प्रतिचयन त्रुटि (ख) अप्रतिचयन त्रुटि

10. मान लीजिए आपकी कक्षा में 10 छात्र है। इनमें से आपको तीन को चुनने हैं, तो इसमें कितने प्रतिदर्श संभव हैं?

11. अपनी कक्षा के 10 छात्रों में से 3 को चुनने के लिए आप लाटरी विधि का उपयोग कैसे करेंगे? चर्चा करें।

12. क्या लाटरी विधि सदैव एक यादृच्छिक प्रतिदर्श देती है? बताएँ।

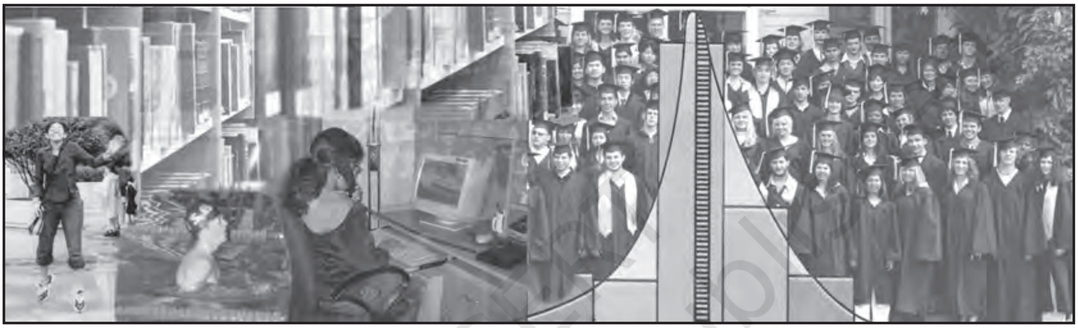
13. यादृच्छिक संख्या सारणी का उपयोग करते हुए, अपनी कक्षा के 10 छात्रों में से 3 छात्रों के चयन के लिए यादृच्छिक प्रतिदर्श की चयन प्रक्रिया की व्याख्या कीजिए।

14. क्या सर्वेक्षणों की अपेक्षा प्रतिदर्श बेहतर परिणाम देते हैं? अपने उत्तर की कारण सहित व्याख्या करें।

© NCERT not to be republished

अध्याय
3

आँकड़ों का संगठन



इस अध्याय को पढ़ने के बाद आप इस योग्य होंगे कि:

- *आगे के सांख्यिकीय विश्लेषण के लिए आँकड़ों का वर्गीकरण कर सकें;*
- *मात्रात्मक एवं गुणात्मक वर्गीकरण के बीच अंतर कर सकें;*
- *बारंबारता वितरण सारणी तैयार कर सकें;*
- *वर्गों के निर्माण की तकनीक जान सकें;*
- *मिलान-चिह्न की विधि से परिचित हो सकें;*
- *एकचर तथा द्विचर बारंबारता वितरण के बीच अंतर कर सकें।*

1. प्रस्तावना

पिछले अध्याय में आपने पढ़ा कि आँकड़ों का संग्रहण कैसे करते हैं। साथ ही, आप जनगणना एवं प्रतिचयन के बीच अंतर को भी जान चुके हैं। इस अध्याय में आप यह सीखेंगे कि जो आँकड़ें आपने संगृहीत किए थे, उन्हें कैसे वर्गीकृत करते हैं। अपरिष्कृत आँकड़ों को वर्गीकृत करने का उद्देश्य उन्हें व्यवस्थित करना है, ताकि उन्हें आसानी से आगे के सांख्यिकीय विश्लेषण के योग्य बनाया जा सके।

क्या आपने कभी स्थानीय कबाड़ी वाले या रद्दी सामान खरीदने वाले को देखा है, जिसे आप अपना पुराना अखबार, टूटे-फूटे घरेलू सामान, खाली-काँच की बोतलें, प्लास्टिक आदि बेचते हैं। वह आपसे इन चीजों को खरीदता है और उन लोगों को बेच देता है जो इनका पुनः चक्रण करते हैं। लेकिन अपनी दुकान में अधिक कबाड़ के इकट्ठे होने से उसे अपना व्यापार चलाने में मुश्किल हो सकती है, अगर वह इन्हें उचित ढंग से व्यवस्थित न करे। वह इस स्थिति को सरल बनाने के लिए विभिन्न कबाड़ों को उपयुक्त समूह में रखता है, अर्थात् उन्हें *वर्गीकृत* करता है। वह पुराने अखबारों को एक साथ रस्सी से बाँध कर रखता है। इसके बाद सभी खाली काँच की बोतलों को एक बोरे

© NCERT not to be republished

में रखता है। वह धातु के सामानों का एक ढेर अपनी दुकान के एक कोने में लगाता है और फिर उनको ‘लोहा’, ‘पीतल’, ‘ताँबा’, ‘एल्यूमिनियम’ आदि वर्गों में छाँट कर रखता है। इस प्रकार से वह अपने कबाड़ को भिन्न वर्गों – ‘अखबार’, ‘प्लास्टिक’, ‘काँच’, ‘धातु’ आदि में विभाजित कर उन्हें व्यवस्थित करता है। जब एक बार उसका सारा कबाड़ व्यवस्थित एवं वर्गीकृत हो जाता है, तब खरीददार की माँग पर, उसे सामग्री विशेष को खोजकर देने में आसानी हो जाती है।

ठीक इसी प्रकार से, जब आप अपने विद्यालय की पुस्तकों को एक विशेष क्रम में रखते हैं, तो उनको संभालना आसान हो जाता है। आप उन्हें विषयों

के अनुसार वर्गीकृत कर सकते हैं, जहाँ प्रत्येक विषय एक समूह या वर्ग बन जाता है। उदाहरणार्थ, जब आपको इतिहास की कोई विशेष पुस्तक की आवश्यकता पड़ती है तो आप को केवल यह करना है कि ‘इतिहास’ समूह में उस पुस्तक को खोजें। अन्यथा आप को अपनी यह विशेष पुस्तक सारी पुस्तकों के ढेर में खोजनी पड़ेगी।

यद्यपि पदार्थों अथवा वस्तुओं का वर्गीकरण बहुमूल्य श्रम और समय को बचाता है, इसे मनमाने तरीके से नहीं किया जाता है। कबाड़ी वाले ने अपने कबाड़ को इस तरह से समूहों में रखा कि प्रत्येक समूह में एक ही प्रकार की चीजें हों। उदाहरण के लिए, उसने ‘काँच’ के समूह में खाली काँच की बोतलें, टूटे खिड़की के काँच तथा टूटे दर्पण आदि रखे। ठीक इसी तरह से जब आपने अपनी इतिहास की पुस्तक को ‘इतिहास’ समूह में वर्गीकृत किया, तो आप उसमें अन्य विषयों की पुस्तकें नहीं रखेंगे। अन्यथा समूह-गठन का पूरा उद्देश्य ही निरर्थक हो जाएगा। *इसलिए, वर्गीकरण का तात्पर्य एक वस्तुओं को समूह या वर्गों में किसी खास आधार पर वर्गीकृत या व्यवस्थित करने से है।*

क्रियात्मक गतिविधि

- अपने स्थानीय डाकघर जायें और देखें कि पत्रों कि छँटाई कैसे की जाती है। क्या आप जानते हैं कि पत्र में पिन कोड का क्या अर्थ है। अपने डाकिए से पूछें।

2. अपरिष्कृत आँकड़े

कबाड़ीवाले के कबाड़ की भाँति, अवर्गीकृत आँकड़े अथवा *अपरिष्कृत आँकड़े* भी अत्यधिक अव्यवस्थित होते हैं। ये प्राय: अति विशाल होते हैं, जिन्हें संभालना कठिन होता है। इनसे सार्थक निष्कर्ष निकालना श्रमसाध्य कार्य है, क्योंकि सांख्यिकीय विधियों का इन पर सरलता से प्रयोग नहीं किया जा सकता। इसलिए इस प्रकार के आँकड़ों का उचित संगठन तथा प्रस्तुतीकरण आवश्यक होता है, ताकि व्यवस्थित रूप से साँख्यिकीय विश्लेषण किया जा सके। अत: आँकड़ों के संग्रह के पश्चात् अगला चरण उन्हें संगठित कर वर्गीकृत रूप में प्रस्तुत करना है।

मान लीजिए, कि आप गणित में छात्रों की प्रगति जानना चाहते हैं और आपने अपने स्कूल के 100 छात्रों के गणित के अंकों के आँकड़े एकत्रित कर लिये

हैं। अगर आप इन्हें एक सारणी में प्रस्तुत करते हैं तो वे संभवत: सारणी 3.1 जैसे प्रतीत हो सकते हैं।

सारणी 3.1

किसी परीक्षा में 100 छात्रों द्वारा गणित में प्राप्त अंक

47	45	10	60	51	56	66	100	49	40
60	59	56	55	62	48	59	55	51	41
42	69	64	66	50	59	57	65	62	50
64	30	37	75	17	56	20	14	55	90
62	51	55	14	25	34	90	49	56	54
70	47	49	82	40	82	60	85	65	66
49	44	64	69	70	48	12	28	55	65
49	40	25	41	71	80	0	56	14	22
66	53	46	70	43	61	59	12	30	35
45	44	57	76	82	39	32	14	90	25

या फिर आप अपने पड़ोस के 50 परिवारों से, भोजन पर उनके मासिक व्यय के आँकड़ों का संग्रह यह जानने के लिए करते हैं कि भोजन पर उनका औसत व्यय कितना है। इस मामले में संगृहीत आँकड़ों को जब आप सारणी में प्रस्तुत करते हैं, तो वे सारणी

3.2 की तरह दिख सकते हैं। सारणी 3.1 तथा सारणी 3.2, दोनों ही आँकड़े अपरिष्कृत अथवा अवर्गीकृत हैं। दोनों ही सारणियों में संख्याओं को किसी भी क्रम में व्यवस्थित नहीं किया गया है। अब अगर आपसे यह पूछा जाए कि सारणी 3.1 में गणित में सर्वोच्च अंक कितने हैं, तब आपको 100 छात्रों के अंकों को आरोही या अवरोही क्रम में व्यवस्थित करना होगा। यह एक बेहद थका देने वाला काम है। यदि आपको 100 छात्रों के स्थान पर 1000 छात्रों के अंक संभालने हों तो यह और भी अधिक थकानेवाला होगा।

सारणी 3.2

खाद्य पर 50 परिवारों के मासिक पारिवारिक व्यय (रु में)

1904	1559	3473	1735	2760
2041	1612	1753	1855	4439
5090	1085	1823	2346	1523
1211	1360	1110	2152	1183
1218	1315	1105	2628	2712
4248	1812	1264	1183	1171
1007	1180	1953	1137	2048
2025	1583	1324	2621	3676
1397	1832	1962	2177	2575
1293	1365	1146	3222	1396

ठीक इसी प्रकार से, सारणी 3.2 में आपके लिए काफी मुश्किल होगा कि 50 परिवारों के खाने पर मासिक व्यय के औसत को पता कर सकें। यही कठिनाई तब कई गुना बढ़ जाएगी यदि यह संख्या बहुत बड़ी हो, जैसे 5000 परिवार। ठीक कबाड़ीवाले की भाँति ही (जब कबाड़ का ढेर बहुत बड़ा और अव्यवस्थित हो तो उसे एक विशेष वस्तु को ढूँढ़ने में बहुत कठिनाई होती है) आपकी भी स्थिति होगी, यदि अपरिष्कृत आँकड़ों का भंडार बहुत बड़ा हो और आप उससे कोई सूचना प्राप्त करना चाहें। इसलिए, संक्षेप में अवर्गीकृत विशाल आँकड़ों से कोई सूचना प्राप्त करना एक बेहद थका देने वाला एवं उबाऊ काम है।

वर्गीकरण के द्वारा अपरिष्कृत आँकड़ों को संक्षिप्त एवं बोधगम्य बनाया जाता है। जब एक प्रकार की विशेषताओं वाले तथ्यों को एक ही वर्ग में रखा जाता है तो वे बिना किसी कठिनाई के ढूँढ़ने, तुलना करने

© NCERT not to be republished

तथा निष्कर्ष निकालने योग्य हो जाते हैं। आपने अध्याय 2 में पढ़ा है कि प्रति दस साल बाद भारत सरकार जनसंख्या की गणना कराती है। सन् 2001 की जनगणना में लगभग 20 करोड़ लोगों से संपर्क किया गया। जनगणना के अपरिष्कृत आँकड़े बहुत विशाल एवं विखंडित होते हैं। उन से कोई भी अर्थपूर्ण निष्कर्ष निकालना असंभव कार्य लगता है। लेकिन जनगणना के यही आँकड़े जब लिंग, शिक्षा, वैवाहिक स्थिति, पेशे आदि के अनुसार वर्गीकृत किये जाते हैं तब भारत की जनसंख्या की प्रकृति एवं संरचना आसानी से समझ में आ जाती है।

अपरिष्कृत आँकड़े चरों के प्रेक्षणों से बने होते हैं। सारणी 3.1 तथा 3.2 में दिए गए अपरिष्कृत आँकड़े विशेष या चर समूह पर किए गए प्रेक्षणों को प्रदर्शित करते हैं। उदाहरण के लिए सारणी 3.1 को देखें जिसमें 100 छात्रों द्वारा गणित में प्राप्त किए गए अंकों को दर्शाया गया है। इन अंकों से हम कैसे अर्थ निकाल सकते हैं? गणित के शिक्षक इन अंकों को देखकर सोच रहे होंगे – मेरे छात्रों ने कैसा किया? कितने असफल रहे? आँकड़ों का वर्गीकरण हमारे उद्देश्यों पर निर्भर करता है इस स्थिति में, शिक्षक गहनतापूर्वक समझने की कोशिश करेंगे – छात्रों ने कैसा किया? संभवतया वह बारंबारता वितरण बनाने का चयन करे। इस पर अगले भाग में विवेचना की जायेगी।

क्रियात्मक गतिविधि

- आप अपने परिवार के एक वर्ष के साप्ताहिक व्यय के आँकड़े संगृहीत कीजिए और उसे एक सारणी में व्यवस्थित कीजिए। देखिए कि उसमें कितने प्रेक्षण हैं। आँकड़ों को मासिक आधार पर व्यवस्थित कीजिए और देखिए कि अब कितने प्रेक्षण हैं।

3. आँकड़ों का वर्गीकरण

किसी वर्गीकरण के वर्ग या समूह कई तरीकों से बनाए जा सकते हैं। आप अपनी पुस्तकों को विषयों–'इतिहास', 'भूगोल', 'गणित', 'विज्ञान' आदि में वर्गीकृत करने के स्थान पर इन्हें वर्णमाला के क्रम में लेखकों के आधार पर वर्गीकृत कर सकते हैं। अथवा, आप इन्हें प्रकाशन– वर्ष के आधार पर भी वर्गीकृत कर सकते हैं। आप उन्हें किस प्रकार से वर्गीकृत करना चाहते हैं, यह आपकी आवश्यकता पर निर्भर करेगा।

ठीक इसी प्रकार से, अपरिष्कृत आँकड़ों को भी विभिन्न तरीकों से वर्गीकृत किया जा सकता है जो आपके अध्ययन के उद्देश्य पर निर्भर करता है। उन्हें समय के अनुसार समूहित किया जा सकता है। इस प्रकार के वर्गीकरण को **कालानुक्रमिक वर्गीकरण** कहते हैं। इस प्रकार के वर्गीकरण में, आँकड़ों को समय के संदर्भ–जैसे वर्ष, तिमाही, मासिक या साप्ताहिक आदि के रूप में, आरोही या अवरोही क्रम में वर्गीकृत किया जा सकता है। निम्नलिखित उदाहरण वर्षों के आधार पर भारत की जनसंख्या के वर्गीकरण को दिखाता है। चर 'जनसंख्या' एक **काल-श्रेणी** है, क्योंकि इसमें विभिन्न वर्षों के मानों की एक श्रेणी चित्रित की गई है।

उदाहरण 1

भारत की जनसंख्या (करोड़ में)

वर्ष	*जनसंख्या (करोड़ में)*
1951	35.7
1961	43.8
1971	54.6
1981	68.4
1991	81.8
2001	102.7
2011	121.0

स्थानिक वर्गीकरण के अंतर्गत आँकड़ों का वर्गीकरण भौगोलिक स्थितियों जैसे कि देश, राज्य,

Reprint 2025-26

शहर, जिला, कस्बा आदि के संदर्भानुसार होता है। उदाहरण 2 में विभिन्न देशों में गेहूँ की उपज दिखाई गई है।

उदाहरण 2

विभिन्न देशों में गेहूँ की उपज (2013)

देश	*गेहूँ की उपज (किग्रा/एकड़)*
कनाडा	3594
चीन	5055
फ्रांस	7254
जर्मनी	7998
भारत	3154
पाकिस्तान	2787

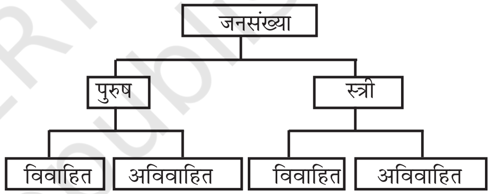
स्रोतः कृषि आँकड़े, भारत सरकार, 2015

क्रियात्मक गतिविधियाँ

- उदाहरण 1 में, उस वर्ष को बताएँ जिसमें भारत की जनसंख्या न्यूनतम और अधिकतम है।
- उदाहरण 2 में, उस देश का पता लगाइये, जिसकी गेहूँ की उपज भारत से थोड़ी अधिक है। यह प्रतिशत में कितनी होगी?
- उदाहरण दो में दिए गए देशों को गेहूँ की उपज के आरोही क्रम में रखिये। ठीक यही अभ्यास उपज को अवरोही क्रम में रखते हुए कीजिए।

कई बार आपका सामना ऐसी विशेषताओं से होता है, जिन्हें मात्रात्मक रूप में व्यक्त नहीं किया जा सकता है। इस प्रकार की विशेषताओं को *'गुण'* कहते हैं। उदाहरण के लिए-राष्ट्रीयता, साक्षरता, धर्म, लिंग, वैवाहिक स्थिति आदि। इन्हें मापा नहीं जा सकता है। इन गुणों को *गुणात्मक* विशेषता की उपस्थिति या अनुपस्थिति के आधार पर वर्गीकृत कर सकते हैं। विशेषताओं पर आधारित आँकड़ों के ऐसे वर्गीकरण को *गुणात्मक वर्गीकरण* कहा जाता है। निम्नलिखित उदाहरण में हम किसी देश की जनसंख्या को गुणात्मक चर 'लिंग' के आधार पर समूहित किया हुआ पाते हैं। इसमें प्रेक्षण स्त्री या पुरुष हो सकता है। इन दो विशेषताओं को आगे वैवाहिक स्थिति के आधार पर वर्गीकृत किया जा सकता है, जैसा कि नीचे दिया गया है

उदाहरण 3



प्रथम चरण में यह वर्गीकरण पहले विशेषता की उपस्थिति या अनुपस्थिति पर आधारित है जैसे कि 'पुरुष' या 'पुरुष नहीं' (स्त्री) है। दूसरे चरण में, प्रत्येक वर्ग 'स्त्री' या 'पुरुष' आगे दूसरी विशेषता की उपस्थिति या अनुपस्थिति के आधार पर विभाजित है, जैसे विवाहित या अविवाहित। ऊँचाई, भार, आयु, आय, छात्रों के अंक आदि विशेषताओं की प्रकृति मात्रात्मक है। जब ऐसी विशिष्टताओं के संगृहीत आँकड़ों को वर्गो में समूहित किया जाता है तो यह वर्गीकरण *मात्रात्मक वर्गीकरण* कहलाता है।

क्रियात्मक गतिविधियाँ

- आस-पास की वस्तुओं को सजीव या निर्जीव के रूप में समूहित किया जा सकता है। क्या यह मात्रात्मक वर्गीकरण है?

© NCERT not to be republished

उदाहरण 4

100 छात्रों के गणित के प्राप्तांकों का बारंबारता वितरण

अंक	*बारंबारता*
0–10	1
10–20	8
20–30	6
30–40	7
40–50	21
50–60	23
60–70	19
70–80	6
80–90	5
90–100	4
योग	*100*

उदाहरण 4 में 100 छात्रों के गणित के प्राप्तांकों का मात्रात्मक वर्गीकरण दिखाया गया है, जिन्हें सारणी 3.1 में बारंबारता वितरण के रूप में दिया गया है।

क्रियात्मक गतिविधियाँ

- उदाहरण 4 की बारंबारता के मानों को कुल बारंबारता के अनुपात में या प्रतिशत में प्रकट कीजिए। ध्यान रहे कि इस प्रकार से प्रकट की गई बारंबारता को सापेक्षिक बारंबारता के रूप में जाना जाता है।
- उदाहरण 4 में किस वर्ग के अंतर्गत आँकड़ों का अधिकतम संकेंद्रण है? इसे कुल प्रेक्षणों के प्रतिशत के रूप में प्रकट कीजिए। किस वर्ग में आँकड़ों का न्यूनतम संकेंद्रण है?

4. चर : संतत और विविक्त

चर की सरल परिभाषा, जिसका आपने पिछले अध्याय में अध्ययन किया था, यह नहीं बतलाती कि यह कैसे परिवर्तित होता है। चरों में अंतर विशेष वर्गीकरण के आधार पर होता है इन्हें सामान्यत: दो वर्गों में वर्गीकृत किया जाता है:

(क) *संतत* तथा

(ख) *विविक्त*

संतत चर का कोई भी संख्यात्मक मान हो सकता है। यह पूर्णांक मान (1, 2, 3, 4 ...), भिन्नात्मक मान (1/2, 2/3, 3/4), तथा वे मान जो यथातथ भिन्न नहीं हैं ($\sqrt{2}$ = 1.414, $\sqrt{3}$ = 1.732, ... , $\sqrt{7}$ = 2.645) हो सकते हैं। उदाहरण के लिए, मान लिजिए कि एक छात्र का कद 90–150 सेमी तक बढ़ता है, तो उसके कद के मान इसके बीच आने वाले सभी मान हो सकते हैं। यह संपूर्ण संख्या वाले मान को भी प्रकट कर सकता है, जैसे कि 90 सेमी, 100 सेमी, 108 सेमी, 150 सेमी। इसके साथ ही यह भिन्नात्मक मान जैसे 90.85 सेमी, 102.34 सेमी, 149.99 सेमी आदि भी हो सकते हैं, जो पूर्णांक नहीं हैं। इस प्रकार 'ऊँचाई' चर किसी भी कल्पित मान को अभिव्यक्त करने में सक्षम है और इसके मानों को अनन्त श्रेणियों में बाँटा जा सकता है। संतत चर के अन्य उदाहरण भार, समय तथा दूरी आदि हैं।

संतत चर के विपरीत *विविक्त चर* केवल निश्चित मान हो सकते हैं। इसके मान केवल परिमित 'उछाल' से बदलते हैं। यह उछाल एक मान से दूसरे मान के बीच होते हैं, परंतु इसके बीच में कोई मान नहीं आता है। उदाहरण के लिए, कोई चर जैसे 'किसी कक्षा में छात्रों की संख्या', भिन्न वर्गों के लिए उन मानों की कल्पना करता है, जिसमें केवल पूर्ण संख्याएँ हों। यह कोई भी भिन्नात्मक मान जैसे 0.5 नहीं हो सकता, क्योंकि 'एक छात्र का आधा' निरर्थक है। इस प्रकार से इसमें 25 एवं 26 के बीच का मान 25.5 नहीं हो सकता है। इसकी अपेक्षा इसका मान या तो 25 होगा या फिर 26। हम देखते हैं कि जब इसका मान 25 से 26 में बदलता है, तो इन दोनों के बीच के भिन्नों को इसमें नहीं लिया जाता है। लेकिन ऐसा नहीं सोचना चाहिए कि किसी विविक्त चर का मान भिन्न में नहीं हो सकता। मान लीजिए कि X एक चर है जिसमें 1/8, 1/16, 1/32, 1/64 ..., जैसे मान हैं तो क्या यह एक विविक्त चर है? हाँ, क्योंकि यद्यपि

© NCERT not to be republished

X के मान भिन्नों में हो सकते हैं, तथापि ये दो सन्निकट भिन्नों के बीच नहीं हो सकते। यह 1/8 से 1/16 में और फिर 1/16 से 1/32 में 'बदलता' है, परंतु 1/8 से 1/16 के बीच या 1/16 से 1/32 के बीच के मान नहीं ले सकता।

क्रियात्मक गतिविधि

- निम्नलिखित चरों का संतत तथा विविक्त में वर्गीकरण करें:
  क्षेत्रफल, आयतन, ताप, पाँसे पर आने वाली संख्या, फसल-उपज, जनसंख्या, वर्षा, सड़क पर कारों की संख्या और आयु।

हमने पहले यह बताया है कि उदाहरण 4 में 100 छात्रों के गणित में प्राप्तांक का बारंबारता वितरण दिया गया है, जैसा कि सारणी 3.1 में दिखाया गया है। यह दिखाता है कि 100 छात्रों के अंकों को वर्गों में कैसे समूहित किया गया है। आपको आश्चर्य होगा कि हमने सारणी 3.1 के अपरिष्कृत आँकड़ों से इसे कैसे प्राप्त किया। लेकिन इस प्रश्न का समाधान प्रस्तुत करने से पहले आपका यह जानना आवश्यक है कि बारंबारता वितरण क्या होता है।

5. बारंबारता वितरण क्या है?

बारंबारता वितरण अपरिष्कृत आँकड़ों को एक मात्रात्मक चर में वर्गीकृत करने का एक सामान्य तरीका है। यह दिखाता है कि किसी चर के भिन्न मान (यहाँ छात्र द्वारा गणित में प्राप्तांक) विभिन्न वर्गों में, अपने अनुरूप वर्गों की बारंबारताओं के साथ कैसे वितरित किए जाते हैं। इस उदाहरण में हमारे पास प्राप्तांकों के 10 वर्ग हैं। 0–10, 10–20, ... 90–100। *वर्ग-बारंबारता* पद का अर्थ है एक विशेष वर्ग में मानों की संख्या। उदाहरण के लिए वर्ग 30–40 में सारणी 3.1 में प्राप्तांकों के 7 मान हैं। ये 30, 37, 34, 30, 35, 39, 32 हैं। इस प्रकार से वर्ग 30–40 की बारंबारता 7 हुई। पर शायद आपको आश्यर्च हो कि 40 का अंक जो अपरिष्कृत आँकड़ों में दो बार आया है, उसे 30–40 वर्ग में शामिल क्यों नहीं किया गया? अगर इसे 30–40 वर्ग की बारंबारता में शामिल किया जाता, तो ये 7 की अपेक्षा 9 होते। यह पहेली तब स्पष्ट हो जाएगी, जब आप इस अध्याय को पर्याप्त धैर्य के साथ सावधानी पूर्वक पढ़ेंगे। इसलिए पढ़ना जारी रखें। आपको स्वयं ही इसका उत्तर प्राप्त हो जाएगा।

बारंबारता वितरण सारणी में प्रत्येक वर्ग, *वर्ग सीमाओं* द्वारा घिरा होता है। वर्ग में ये सीमाएँ दो छोरों पर होती हैं। इसमें न्यूनतम मान को *निम्नवर्ग सीमा* तथा उच्चतम मान को *उच्च वर्ग सीमा* कहते हैं। उदाहरण के लिए वर्ग 60–70 में वर्ग सीमाएँ 60 एवं 70 हैं। इसकी निम्न वर्ग सीमा 60 और उच्च वर्ग सीमा 70 है। *वर्ग मध्यांतर* या अंतराल या *वर्ग विस्तार* उच्च वर्ग सीमा तथा निम्न वर्ग सीमा के बीच का अंतर है। वर्ग 60–70 के लिए वर्ग अंतराल 10 है, (उच्च वर्ग सीमा में से निम्न वर्ग सीमा को *घटाकर*)।

वर्ग मध्यबिन्दु अथवा *वर्ग चिह्न* किसी वर्ग का मध्य-मान है। यह वर्ग की निम्न वर्ग सीमा तथा उच्च वर्ग सीमा के बीच होता है। इसे निम्नलिखित तरीके से पता किया जा सकता है:

वर्ग मध्य बिन्दु या *वर्ग चिह्न = (उच्च वर्ग सीमा + निम्न वर्ग सीमा)/2 ... ... (1)*

प्रत्येक वर्ग का वर्ग चिह्न या वर्ग मध्य-बिन्दु एक वर्ग के प्रतिनिधित्व के लिए प्रयुक्त किया जाता है। एक बार जब अपरिष्कृत आँकड़ों को वर्गों में समूहित कर दिया जाता है, तब आगे की गणनाओं में व्यष्टि प्रेक्षणों का प्रयोग नहीं किए जाता है बल्कि इसकी जगह वर्ग चिह्न प्रयुक्त किया जाता है।

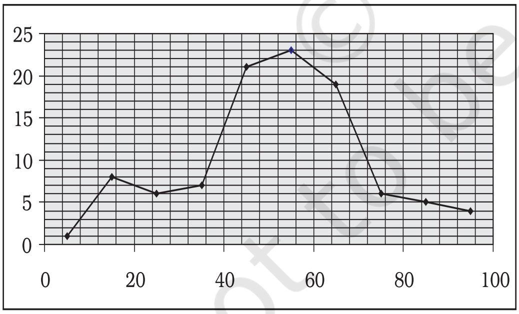
© NCERT not to be republished

सारणी 3.3

निम्न वर्ग सीमा, उच्च वर्ग सीमा तथा वर्ग चिह्न

वर्ग	*बारंबारता*	*निम्नवर्ग सीमा*	*उच्चवर्ग सीमा*	*वर्ग चिह्न*
0–10	1	0	10	5
10–20	8	10	20	15
20–30	6	20	30	25
30–40	7	30	40	35
40–50	21	40	50	45
50–60	23	50	60	55
60–70	19	60	70	65
70–80	6	70	80	75
80–90	5	80	90	85
90–100	4	90	100	95

बारंबारता वक्र किसी बारंबारता वितरण का आलेखीय प्रस्तुतीकरण है। चित्र 3.1 के अंतर्गत उपरोक्त उदाहरण में दिए गए आँकड़ों का आरेखी प्रस्तुतीकरण दिया गया है। बारंबारता वक्र प्राप्त करने के लिए, हम वर्ग चिह्न को एक्स (x) अक्ष पर तथा बारंबारता को वाई (y) अक्ष पर आलेखित करते हैं।



चित्र 3.1 *आँकड़ों के बारंबारता वितरण का आरेखी प्रस्तुतीकरण।*

बारंबारता वितरण कैसे तैयार करें?

बारंबारता वितरण तैयार करते समय हमें निम्न पाँच प्रश्नों की व्याख्या पर ध्यान देने की आवश्यकता है:

1. वर्ग अंतराल समान आकार के हों या असमान आकार के?
2. हमें कितने वर्ग रखने चाहिए?
3. प्रत्येक वर्ग का आकार क्या हो?
4. वर्ग सीमाओं का निर्धारण कैसे किया जाय?
5. प्रत्येक वर्ग के लिए बारंबारता कैसे प्राप्त की जाय?

वर्ग अंतराल, समान अंतराल के हों या असमान अंतराल के?

दो परिस्थितियों में असमान आकार के वर्ग अंतरालों का प्रयोग किया जाता है। पहली, जब हमारे पास आय तथा ऐसे ही चरों के आँकड़े हों, जहाँ परास काफी अधिक होता है। उदाहरण के लिए, दैनिक आय लगभग शून्य से लेकर कई सौ करोड़ रुपये तक हो सकती है। ऐसी स्थिति में, समान वर्ग अंतराल उपयुक्त नहीं है, क्योंकि (i) यदि वर्ग अंतराल छोटे तथा समान आकार के होंगे, तो वर्गों की संख्या बहुत अधिक हो जाएगी। (ii) यदि वर्ग अंतराल अधिक है, तो आय के बहुत कम या बहुत अधिक स्तरों पर जानकारी छिपी हुई रह जाएगी।

दूसरी, यदि मानों की एक बहुत बड़ी संख्या परास के एक छोटे से भाग में केंद्रित होती है, तो समान वर्ग अंतराल से कई मानों की सूचना प्राप्त नहीं हो पाएगी।

अन्य सभी स्थितियों में, आवृत्ति-वितरण में समान आकार के वर्ग अंतरालों का प्रयोग होता है।

वर्गों की संख्या कितनी होनी चाहिए?

वर्गों की संख्या सामान्यत: 6 तथा 15 के बीच होती है। यदि हमारे वर्ग अंतराल समान आकार के हों, तो वर्गों की संख्या, परास (चर के अधिकतम तथा न्यूनतम मान में अंतर) को वर्ग अंतराल से भाग देने पर प्राप्त की जा सकती है।

क्रियात्मक गतिविधियाँ

निम्नलिखित का परास ज्ञात करें:
- उदाहरण 1 में भारत की जनसंख्या।
- उदाहरण 2 में गेहूँ की उपज।

प्रत्येक वर्ग का आकार क्या होना चाहिए?

इस प्रश्न का उत्तर पहले के प्रश्न के उत्तर पर निर्भर करता है। समीकरण (2) प्रकट करती है कि एक बार वर्ग अंतराल को तय करने पर चर के दिए गए परास से हम वर्गों की संख्या निर्धारित कर सकते हैं। ठीक इसी प्रकार से हम वर्ग अंतराल निर्धारित कर सकते हैं, जब एक बार हम वर्गों की संख्या तय कर लेते हैं। इस तरह हम पाते हैं कि ये दोनों निर्णय एक दूसरे से जुड़े हुए हैं। पहले का निर्णय लिए बिना हम दूसरे पर निर्णय नहीं ले सकते।

उदाहरण 4 में, हमारे पास वर्गों की संख्या 10 है तथा परास का दिया गया मान 100 है, तब वर्ग–अंतराल स्वत: ही (समानता 2 के द्वारा) 10 है। ध्यान दें कि वर्तमान संदर्भ में हमने वह वर्ग अंतराल चुना है, जिनका परिमाण समान है। तथापि हम ऐसा वर्ग अंतराल चुन सकते हैं जिसका परिमाण समान न हो, तब ऐसे मामले में वर्गों की चौड़ाई असमान होगी।

हमें वर्ग सीमाएँ कैसे निर्धारित करनी चाहिए?

वर्ग सीमाएँ निश्चित तथा स्पष्ट रूप से होनी चाहिए। सामान्यत: मुक्तोत्तर वर्ग, जैसे– '70 तथा अधिक' या '10 से कम' वांछनीय नहीं होते। निम्न तथा उच्च वर्ग सीमाओं का निर्धारण इस प्रकार से किया जाना चाहिए कि प्रत्येक वर्ग की आवृत्तियों की प्रवृत्ति वर्ग अंतराल के मध्य में संकेंद्रण की हो। वर्ग अंतराल दो प्रकार के होते हैं–

1. समावेशी वर्ग अंतराल: इस स्थिति में, वर्ग की निम्न तथा उच्च सीमाओं के मूल्य वाले मानों को उस वर्ग की आवृत्ति में शामिल किया जाता है।
2. अपवर्जी वर्ग अंतराल: इस स्थिति में, वर्ग की निम्न तथा उच्च सीमाओं के मूल्य वाली मदों को उस वर्ग की आवृत्ति में शामिल नहीं किया जाता।

असतत चरों की स्थिति में, अपवर्जी तथा समावेशी, दोनों प्रकार के वर्ग अंतरालों का प्रयोग किया जा सकता है।

सतत चरों की स्थिति में, समावेशी वर्ग अंतरालों का प्रयोग बहुधा किया जाता है।

उदाहरण

मान लीजिए, हमारे पास एक परीक्षा में विद्यार्थियों द्वारा प्राप्तांकों के आँकड़े हैं तथा सभी प्राप्तांक पूर्णांक हैं (भिन्नात्मक अंकों की अनुमति नहीं है)। मान लीजिए, विद्यार्थियों द्वारा प्राप्तांक 0 से 100 के बीच हैं।

यह असतत चरों की स्थिति है, क्योंकि भिन्नात्मक अंकों की अनुमति नहीं है। इस स्थिति में, यदि हम समान आकार वाले वर्ग अंतरालों का उपयोग करते हैं तथा 10 वर्ग अंतरालों का प्रयोग करते हैं, तो वर्ग अंतरालों के निम्न रूप हो सकते हैं–

वर्ग अंतराल का समावेशी रूप

0–10
11–20
21–30
–
–
91–100

वर्ग अंतराल का अपवर्जी रूप

0–10
10–20
20–30
–
–
90–100

अपवर्जी वर्ग अंतराल की स्थिति में, हमें यह अग्रिम रूप से निर्धारित करना होता है कि वर्ग सीमा

© NCERT not to be republished

के मान के बराबर किसी चर का मान होने पर क्या करना है। उदाहरण के लिए, हम यह निर्णय कर सकते हैं कि 10, 30 आदि मानों को क्रमशः वर्ग अंतराल ''0 से 10'' तथा ''20 से 30'' में रखा जाए। इस स्थिति में वर्ग की निचली सीमा को वर्ग अंतराल में शामिल नहीं किया जाता।

या फिर हम 10, 30 आदि मानों को क्रमशः वर्ग अंतराल ''10 से 20'' तथा ''30 से 40'' में रख सकते हैं। इस स्थिति में वर्ग की उच्च सीमा को वर्ग अंतराल में शामिल नहीं किया जाता।

सतत चर के उदाहरण

मान लें कि हमारे पास किसी चर के आँकड़े उपलब्ध हों, जैसे कद (से.मी.) या वज़न (कि.ग्रा.)। यह आँकड़ा सतत प्रकार का है। ऐसी स्थितियों में वर्ग अंतराल निम्नलिखित प्रकार से दर्शाया जा सकता है–

30 कि.ग्रा.– 39.999...कि.ग्रा.

40 कि.ग्रा.– 49.999...कि.ग्रा.

50 कि.ग्रा.– 59.999...कि.ग्रा. आदि।

इन वर्ग अंतरालों को निम्नलिखित प्रकार से समझा जा सकता है–

30 कि.ग्रा. और अधिक तथा 40 कि.ग्रा. से कम

40 कि.ग्रा. और अधिक तथा 50 कि.ग्रा. से कम

50 कि.ग्रा. और अधिक तथा 60 कि.ग्रा. से कम आदि।

सारणी 3.4

एक कंपनी के 550 कर्मचारियों की आय का बारंबारता वितरण

आय (रु में)	*कर्मचारियों की संख्या*
800–899	50
900–999	100
1000–1099	200
1100–1199	150
1200–1299	40
1300–1399	10
योग	*550*

वर्ग अंतराल में समायोजन

सारणी 3.4 में समावेशी विधि के सूक्ष्म अध्ययन से पता चलता है कि यद्यपि चर 'आय' एक संतत चर है, तथापि जब वर्गों को बनाया जाता है तो संततता नहीं रहती। हम एक वर्ग की उच्च सीमा तथा अगले वर्ग की निम्न सीमा में 'अंतर' या असंततता पाते हैं। उदाहरण के लिए, पहले वर्ग की उच्च सीमा 899 और दूसरे वर्ग की निम्न सीमा 900 के बीच हम 1 (एक) का 'अंतर' पाते हैं। तब हम आँकड़ों के वर्गीकरण में चर की संततता को कैसे सुनिश्चित करते हैं? इसे वर्ग अंतराल के बीच समायोजन करके किया जाता है। समायोजन निम्नलिखित तरीके से किया गया है।

1. द्वितीय वर्ग की निम्न सीमा और प्रथम वर्ग की उच्च सीमा के बीच अंतर पता करें। उदाहरण के लिए, सारणी 3.4 में द्वितीय वर्ग की निम्न सीमा 900 और प्रथम वर्ग की उच्च सीमा 899 के बीच अंतर 1 है (अर्थात 900 – 899 = 1)।
2. प्राप्त किए गए अंतर (1) को 2 से विभाजित करें (अर्थात 1/2 = 0.5)।
3. सभी वर्गों की निम्न सीमाओं से (2) में प्राप्त किए गए मान को घटाइए (निम्न वर्ग सीमा – 0.5)।
4. सभी वर्गों की उच्च सीमा में (2) में प्राप्त किए गए मान को जोड़िए (उच्च वर्ग सीमा + 0.5)।

समायोजन के पश्चात्, जिससे बारंबारता वितरण में आँकड़ों की संततता की पुनः प्राप्ति होती है, सारणी 3.4 संशोधित होकर सारणी 3.5 बन जाती है।

वर्ग सीमाओं में समायोजन के पश्चात्, समानता (1) जोकि वर्ग चिह्न का मान निर्धारित करती है, निम्नलिखित प्रकार से संशोधित हो जाएगी:

समायोजित वर्ग चिह्न = (समायोजित उच्च वर्ग सीमा + समायोजित निम्न वर्ग सीमा)/2

सारिणी 3.5

एक कंपनी के 550 कर्मचारियों की आय का बारंबारता वितरण

आय (रु में)	*कर्मचारियों की संख्या*
799.5–899.5	50
899.5–999.5	100
999.5–1099.5	200
1099.5–1199.5	150
1199.5–1299.5	40
1299.5–1399.5	10
योग	*550*

हमें प्रत्येक वर्ग की बारंबारता कैसे प्राप्त करनी चाहिए

साधारण शब्दों में, *एक प्रेक्षण की बारंबारता का अर्थ है कि अपरिष्कृत आँकड़ों में कितनी बार वह प्रेक्षण प्रकट होता है।* सारणी 3.1 में, हमने देखा कि 40 का मान तीन बार आया है, जबकि 0 और 10 का मान एक बार, 49 का मान 5 बार और ऐसे ही अन्य मान आये हैं। इस प्रकार से 40 की बारंबारता 3, 0 की 1, 10 की 1, 49 की 5 तथा ऐसे ही। लेकिन जब आँकड़े वर्गों में समूहित कर दिए जाते हैं, जैसा कि उदाहरण 3 में किया गया है, तो किसी वर्ग की बारंबारता से तात्पर्य उस वर्ग के मानों की संख्याओं से है। वर्ग-बारंबारताओं की गिनती विशेष वर्ग के सामने मिलान चिह्नों को लगाकर की जाती है।

मिलान चिह्न अंकन द्वारा वर्ग बारंबारता को ज्ञात करना

मिलान चिह्न (/) किसी वर्ग के प्रत्येक छात्र के सामने लगाया जाता है, जिसके प्राप्तांक उस वर्ग में शामिल हैं। उदाहरण के लिए, यदि किसी छात्र का प्राप्तांक 57 है तो उस छात्र के लिए वर्ग 50–60 में एक मिलान चिह्न (/) लगाया जाता है। यदि प्राप्तांक 71 हैं तो मिलान चिह्न (/) को वर्ग 70–80 में लगाया जाता है। यदि कोई 40 अंक प्राप्त करता है तो उसके लिए मिलान चिह्न वर्ग 40–50 में लगाया जाता है। सारणी 3.1 के 100 छात्रों के गणित में प्राप्तांकों के मिलान चिह्नों को सारणी 3.6 में दिखाया गया है।

मिलान चिह्नों का परिकलन तब आसान हो जाता है जब 4 चिह्न खड़े (////) लगाए जाते हैं और पाँचवाँ चिह्न सबको काटता हुआ तिरछा लगाया जाता है, जैसे (~~////~~)। मिलान चिह्नों की गणना पाँच के समूह में की जाती है, इसलिए यदि किसी वर्ग में 16 मिलान चिह्न हैं तो उन्हें इस प्रकार से ~~////~~ ~~////~~ ~~////~~ / लिखते हैं, ताकि परिकलन में सुविधा रहे। इसलिए एक वर्ग की बारंबारता उतनी ही होगी, जितनी उस वर्ग में मिलान चिह्नों की संख्या।

सूचना की हानि (Loss of Information)

बारंबारता वितरण के रूप में आँकड़ों के वर्गीकरण में एक अंतर्निहित दोष पाया जाता है। यह अपरिष्कृत आँकड़ों का सारांश प्रस्तुत कर उन्हें संक्षिप्त एवं बोध गम्य तो बनाता है, परंतु इसमें वे विस्तृत विवरण नहीं प्रकट हो पाते जो अपरिष्कृत आँकड़ों में पाए जाते हैं। यद्यपि अपरिष्कृत आँकड़ों को वर्गीकृत करने में सूचना की क्षति होती है, तथापि आँकड़ों को वर्गीकरण द्वारा संक्षिप्त करने पर पर्याप्त जानकारी मिल जाती है। एक बार जब आँकड़ों को वर्गों में समूहित कर दिया जाता है तब व्यष्टि प्रेक्षणों का आगे सांख्यकीय परिकलनों में कोई महत्व नहीं होता। उदाहरण 4 में, वर्ग 20–30 के अंतर्गत 6 प्रेक्षण 25, 25, 20, 22, 25 एवं 28 हैं। इसलिए जब इन आँकड़ों को बारंबारता वितरण में वर्ग 20–30 में समूहित कर दिया जाता है, तब यह बारंबारता वितरण उस वर्ग की बारंबारता (जैसे 6) को दिखाता है, न कि उनके वास्तविक मानों को। *इस वर्ग के सभी मानों को उस वर्ग के वर्ग-अंतराल के मध्य मान या वर्ग चिह्न के बराबर माना जाता है (अर्थात् 25)। आगे की सांख्यिकीय परिकलनों के लिए वर्ग चिह्न के मान को आधार बनाया जाता है, न*

© NCERT not to be republished

कि उस वर्ग के प्रेक्षणों के मान को। यही बात सभी वर्गों के लिए सत्य है। इस प्रकार प्रेक्षणों के वास्तविक मान के स्थान पर वर्ग चिह्नों के प्रयोग को सांख्यिकीय विधियों में शामिल करने पर पर्याप्त मात्रा में सूचनाओं की क्षति होती है। यद्यपि अपरिष्कृत आँकड़ों का जैसा कि नीचे इससे अधिक दिखाया गया है, अधिक अर्थपूर्ण लगता है।

असमान वर्गों में बारंबारता वितरण

अब तक आप समान वर्ग अंतराल के बारंबारता वितरण से परिचित हो चुके हैं। आप जान गए हैं कि इन्हें अपरिष्कृत आँकड़ों से कैसे गठित किया जाता है। लेकिन कुछ मामलों में असमान वर्ग अंतराल के साथ बारंबारता वितरण अधिक उपयुक्त होता है। यदि आप उदाहरण 4 के बारंबारता वितरण की *सारणी 3.6* को देखें, तो आप पायेंगे कि अधिकांश प्रेक्षण वर्ग 40−50, 50−60 तथा 60−70 में संकेंद्रित हैं। उनकी बारंबारताएँ क्रमशः 21, 23 एवं 19 हैं। इसका अर्थ है कि 100 छात्रों में से 63 (21 + 23 + 19) प्रेक्षण इन वर्गों में सकेंद्रित हैं। इस प्रकार 63 प्रतिशत आँकड़े 40-70 के बीच समाहित है और आँकड़ों का शेष 37 प्रतिशत 0−10, 10−20, 20−30, 30−40 तथा 70−80, 80−90 एवं 90−100 वर्गों में हैं। इन वर्गों में प्रेक्षणों का विरल घनत्व है। आप यह भी देखेंगे कि इन वर्गों के प्रेक्षणों में अन्य वर्गों की अपेक्षा उनके अपने वर्गों के वर्ग-चिह्नों से अधिक विचलन है। लेकिन यदि वर्गो का गठन इस प्रकार से करना हो कि वर्ग चिह्न, जहाँ तक संभव है, उस मान के बराबर हो जाए, जिसके आस-पास उस वर्ग के प्रेक्षणों के संकेंद्रण की प्रवृत्ति होती है, तो असमान वर्ग अंतराल अधिक उपयुक्त होता है।

असमान वर्गों के रूप में, सारणी 3.7 में सारणी 3.6 के उसी बारंबारता वितरण को दिखाया गया है। 40−50, 50−60 तथा 60−70 के प्रत्येक वर्ग को दो भागों में विभाजित किया गया है। वर्ग 40−50 को

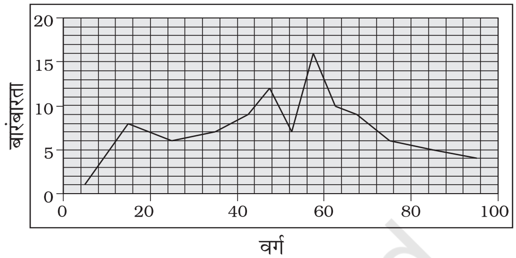
सारणी 3.6

गणित में 100 छात्रों के प्राप्तांको के मिलान चिह्न

वर्ग	*प्रेक्षण*	*मिलान चिह्न*	*बारंबारता*	*वर्ग चिह्न*
0−10	0	/	1	5
10−20	10, 14, 17, 12, 14, 12, 14, 14	~~////~~ ///	8	15
20−30	25, 25, 20, 22, 25, 28	~~////~~ /	6	25
30−40	30, 37, 34, 39, 32, 30, 35,	~~////~~ //	7	35
40−50	47, 42, 49, 49, 45, 45, 47, 44, 40, 44, 49, 46, 41, 40, 43, 48, 48, 49, 49, 40, 41	~~////~~ ~~////~~ ~~////~~ ~~////~~ /	21	45
50−60	59, 51, 53, 56, 55, 57, 55, 51, 50, 56, 59, 56, 59, 57, 59, 55, 56, 51, 55, 56, 55, 50, 54	~~////~~ ~~////~~ ~~////~~ ~~////~~ ///	23	55
60−70	60, 64, 62, 66, 69, 64, 64, 60, 66, 69, 62, 61, 66, 60, 65, 62, 65, 66, 65	~~////~~ ~~////~~ ~~////~~ ////	19	65
70−80	70, 75, 70, 76, 70, 71	~~////~~ /	6	75
80−90	82, 82, 82, 80, 85	~~////~~	5	85
90−100	90, 100, 90, 90	////	4	95
	योग		*100*	

अब 40–45 तथा 45–50 में बाँटा गया है। वर्ग 50–60 को 50–55 और 55–60 में बाँटा गया है तथा वर्ग 60–70 को 60–65 तथा 65–70 में बाँटा गया है। अब नए वर्ग 40–45, 45–50, 50–55, 55–60, 60–65 तथा 65–70 हैं जिनमें वर्ग अंतराल 5 है। बाकी अन्य वर्गों 0–10, 10–20, 20–30, 30–40, तथा 70–80, 80–90, 90–100 में ठीक वही पूर्ववत वर्ग अंतराल 10 है। इस सारणी का अंतिम स्तंभ इन वर्गों के नये वर्ग चिह्नों को प्रदर्शित कर रहा है। सारणी 3.6 के पुराने वर्ग चिह्नों से उनकी तुलना करें। ध्यान दें कि इन वर्गों के प्रेक्षणों में नये वर्ग चिह्न मानों की अपेक्षा पुराने वर्ग चिह्न मानों से विचलन अधिक है। इस प्रकार से नये वर्ग चिह्न मान, इन वर्गों के आँकड़ों का पुराने मान की अपेक्षा बेहतर प्रतिनिधित्व करते हैं।

चित्र 3.2 में, सारणी 3.7 के बारंबारता वितरण के बारंबारता वक्र को दिखाया गया है। इसमें सारणी के वर्ग चिह्नों को X-अक्ष पर तथा बारंबारताओं को y-अक्ष पर आलेखित किया गया है।



3.2 बारंबारता वक्र

क्रियाकलाप

- यदि आप चित्र 3.2 के साथ चित्र 3.1 की तुलना करते हैं तो आप क्या देखते हैं? क्या आपने इनके बीच कोई अंतर पाया? क्या आप उस अंतर की व्याख्या कर सकते हैं?

सारणी 3.7

असमान वर्गों में बारंबारता वितरण

वर्ग	*प्रेक्षण*	*बारंबारता*	*वर्ग चिह्न*
0–10	0	1	5
10–20	10, 14, 17, 12, 14, 12, 14, 14	8	15
20–30	25, 25, 20, 22, 25, 28	6	25
30–40	30, 37, 34, 39, 32, 30, 35	7	35
40–45	42, 44, 40, 44, 41, 40, 43, 40, 41	9	42.5
45–50	47, 49, 49, 45, 45, 47, 49, 46, 48, 48, 49, 49	12	47.5
50–55	51, 53, 51, 50, 51, 50, 54	7	52.5
55–60	59, 56, 55, 57, 55, 56, 59, 56, 59, 57, 59, 55, 56, 55, 56, 55	16	57.5
60–65	60, 64, 62, 64, 64, 60, 62, 61, 60, 62	10	62.5
65–70	66, 69, 66, 69, 66, 65, 65, 66, 65	9	67.5
70–80	70, 75, 70, 76, 70, 71	6	75
80–90	82, 82, 82, 80, 85	5	85
90–100	90, 100, 90, 90	4	95
	योग	*100*	

बारंबारता सरणी (Frequency Array)

अब तक हमने गणित में 100 छात्रों द्वारा प्राप्त किए गए प्रतिशत अंकों के उदाहरण का प्रयोग करते हुए संतत चर के लिए आँकड़ों के वर्गीकरण पर चर्चा की है। विविक्त चर के लिए, आँकड़ों का वर्गीकरण *बारंबारता सरणी* के नाम से जाना जाता है। चूँकि एक विविक्त चर मानों को धारण करता है न कि दो पूर्णाकों के बीच माध्यमिक भिन्नीय मानों को, अतः हम ऐसी बारंबारता रखते हैं जोकि अपने पूर्णांक मानों से संगत हों।

सारणी 3.8
परिवारों के आकार की बारंबारता सारणी

परिवार का आकार	*परिवारों की संख्या*
1	5
2	15
3	25
4	35
5	10
6	5
7	3
8	2
योग	*100*

सारणी 3.8 में दिया गया उदाहरण बारंबारता सरणी को प्रदर्शित करता है। इस सारणी में चर 'परिवार का आकार' एक विविक्त चर है जो सारणी में दिखाए गए पूर्णाकों को ही धारण करता है।

6. द्विचर बारंबारता वितरण

बहुत बार, जब हम किसी जनसंख्या में से एक प्रतिदर्श लेते हैं, तो हम प्रतिदर्श के हर अवयव से एक से अधिक प्रकार की सूचना संगृहीत करते हैं। उदाहरण के लिए, मान लीजिए कि हमने एक शहर की कंपनियों की सूची में से 20 कंपनियों का एक प्रतिदर्श लिया है। मान लीजिए कि हम प्रत्येक कंपनी से P बिक्री तथा विज्ञापनों पर किए गए व्यय की जानकारी संगृहीत करते हैं। इस स्थिति में, हमारे पास प्रतिदर्श के द्विचर आँकड़े हैं। इस तरह के द्विचर आँकड़ों को द्विचर बारंबारता वितरण द्वारा संक्षिप्त रूप में दर्शाया जा सकता है।

एक द्विचर बारंबारता वितरण को दो चरों के बारंबारता वितरण के रूप में परिभाषित किया जा सकता है।

सारणी 3.9, 20 कंपनियों के दो चर-बिक्री एवं विज्ञापन व्यय (लाख रु में) के बारंबारता वितरण को प्रदर्शित कर रही है। यहाँ पर बिक्री मानों को भिन्न स्तंभों में तथा विज्ञापन व्यय के मानों को भिन्न पंक्तियों में वर्णित किया गया है। प्रत्येक प्रकोष्ठ संतत पंक्ति एवं स्तंभ के मान की बारंबारता दिखाता है। उदाहरण के लिए, यहाँ पर तीन फर्म ऐसी हैं, जिनकी बिक्री रु 135–145 लाख रु के

सारणी 3.9
20 कंपनियों की बिक्री (लाख रु में) एवं विज्ञापन व्यय (हजार रु में) का द्विचर बारंबारता वितरण

	115–125	*125–135*	*135–145*	*145–155*	*155–165*	*165–175*	*योग*
62–64	2	1					3
64–66	1		3				4
66–68	1	1	2	1			5
68–70		2		2			4
70–72		1	1		1	1	4
योग	*4*	*5*	*6*	*3*	*1*	*1*	*20*

© NCERT not to be republished

बीच है और उनका विज्ञापन व्यय 64–66 हजार रु के बीच है। द्विचर वितरण के बारे में अध्याय 8 'सहसंबंध' में अध्ययन किया जाएगा।

7. सारांश

प्राथमिक या द्वितीयक स्रोतों से संगृहीत किए गए आँकड़े अपरिष्कृत या अवर्गीकृत होते हैं। जब एक बार आँकड़े संगृहित हो जाएँ तो अगला चरण आगे के सांख्यिकीय विश्लेषण के लिए आँकड़ों का वर्गीकरण करना है। वर्गीकरण से आँकड़ों में क्रमबद्धता आ जाती है। यह अध्याय आप को यह जानने के योग्य बनाता है कि आँकड़ों को बारंबारता वितरण के माध्यम से बोधगम्य तरीके से किस प्रकार वर्गीकृत किया जाता है। एक बार जब आप वर्गीकरण की तकनीकों को जान जाते हैं तो आपके लिए यह आसान होगा कि आप संतत तथा विविक्त दोनों चरों के लिए ही बारंबारता वितरण की रचना कर सकें।

पुनरावर्तन

- वर्गीकरण अपरिष्कृत आँकड़ों को क्रमबद्धता प्रदान करता है।
- बारंबारता वितरण यह प्रदर्शित करता है कि किसी चर के विभिन्न मान, संगत वर्ग–बारंबारताओं सहित, किस प्रकार विभिन्न वर्गों में वितरित किए जाते हैं।
- अपवर्जी विधि के अंतर्गत उच्च वर्ग सीमा को छोड़ा तथा निम्नवर्ग सीमा को शामिल किया जाता है।
- समावेशी विधि में निम्नवर्ग सीमा तथा उच्च वर्ग सीमा, दोनों को ही शामिल किया जाता है।
- बारंबारता वितरण में, आगे के सांख्यिकीय परिकलन केवल वर्ग चिह्न मान पर आधारित होते हैं, न कि प्रेक्षणों के मान पर।
- वर्गों को इस प्रकार से बनाया जाना चाहिए कि, जहाँ तक संभव हो सके, प्रत्येक वर्ग का वर्ग चिह्न उस मान के अधिक से अधिक निकटतम हो, जिस मान के आस-पास, किसी वर्ग के प्रेक्षणों की संकेन्द्रण की प्रवृत्ति हो।

© NCERT not to be republished

अभ्यास

1. निम्नलिखित में से कौन सा विकल्प सही है?
   - एक वर्ग मध्यबिन्दु बराबर है:
     (क) उच्च वर्ग सीमा तथा निम्न वर्ग सीमा के औसत के।
     (ख) उच्च वर्ग सीमा तथा निम्न वर्ग सीमा के गुणनफल के।
     (ग) उच्च वर्ग सीमा तथा निम्न वर्ग सीमा के अनुपात के।
     (घ) उपरोक्त में से कोई नहीं।
   - दो चरों के बारंबारता वितरण को इस नाम से जानते हैं:
     (क) एक विचर वितरण
     (ख) द्विचर वितरण
     (ग) बहुचर वितरण
     (घ) उपरोक्त में से कोई नहीं
   - वर्गीकृत आँकड़ों में साँख्यिकीय परिकलन आधारित होता है:
     (क) प्रेक्षणों के वास्तविक मानों पर
     (ख) उच्च वर्ग सीमाओं पर
     (ग) निम्न वर्ग सीमाओं पर
     (घ) वर्ग के मध्यबिन्दुओं पर
   - अपवर्जी विधि के अंतर्गत
     (क) किसी वर्ग की उच्च वर्ग सीमा को वर्ग अंतराल में समावेशित नहीं करते।
     (ख) किसी वर्ग की उच्च वर्ग सीमा को वर्ग अंतराल में समावेशित करते हैं।
     (ग) किसी वर्ग की निम्न वर्ग सीमा को वर्ग अंतराल में समावेशित नहीं करते।
     (घ) किसी वर्ग की निम्न वर्ग सीमा को वर्ग अंतराल में समावेशित करते हैं।
   - परास का अर्थ है
     (क) अधिकतम एवं न्यूनतम प्रेक्षणों के बीच अंतर
     (ख) न्यूनतम एवं अधिकतम प्रेक्षणों के बीच अंतर
     (ग) अधिकतम एवं न्यूनतम प्रेक्षणों का औसत
     (घ) अधिकतम एवं न्यूनतम प्रेक्षणों का अनुपात
2. वस्तुओं को वर्गीकृत करने में क्या कोई लाभ हो सकता है? अपने दैनिक जीवन से एक उदाहरण देकर व्याख्या कीजिए।
3. चर क्या है? एक संतत तथा विविक्त चर के बीच भेद कीजिए।
4. आँकड़ों के वर्गीकरण में प्रयुक्त अपवर्जी तथा समावेशी विधियों की व्याख्या कीजिए।

© NCERT not to be republished

5. सारणी 3.2 के आँकड़ों का प्रयोग करें, जो 50 परिवारों के भोजन पर मासिक व्यय (रु में) को दिखलाती है, और
   (क) भोजन पर मासिक परिवारिक व्यय का प्रसार ज्ञात कीजिए।
   (ख) परास को वर्ग अंतराल की उचित संख्याओं में विभाजित करें तथा व्यय का बारंबारता वितरण प्राप्त करें।
   - उन परिवारों की संख्या पता कीजिए जिनका भोजन पर मासिक व्यय
     (क) 2000/- रु से कम है
     (ख) 3000/- रु से अधिक है
     (ग) 1500/- रु और 2500/- रु के बीच है
6. एक शहर में, यह जानने हेतु 45 परिवारों का सर्वेक्षण किया गया कि वे अपने घरों में कितनी संख्या में सेल फोनों का इस्तेमाल करते हैं। नीचे दिए गए उनके उत्तरों के आधार पर एक बारंबारता सरणी तैयार कीजिए।

1	3	2	2	2	2	1	2	1	2	2	3	3	3	3
3	3	2	3	2	2	6	1	6	2	1	5	1	5	3
2	4	2	7	4	2	4	3	4	2	0	3	1	4	3

7. वर्गीकृत आँकड़ों में 'सूचना की क्षति' का क्या अर्थ है?
8. क्या आप इस बात से सहमत है कि अपरिष्कृत आँकड़ों की अपेक्षा वर्गीकृत आँकड़े बेहतर होते हैं?
9. एक-विचर एवं द्विचर बारंबारता वितरण के बीच अंतर बताइए?
10. निम्नलिखित आँकडों के आधार पर 7 का वर्ग अंतराल लेकर समावेशी विधि द्वारा एक बारंबारता वितरण तैयार कीजिए।

28	17	15	22	29	21	23	27	18	12	7	2	9	4
1	8	3	10	5	20	16	12	8	4	33	27	21	15
3	36	27	18	9	2	4	6	32	31	29	18	14	13
15	11	9	7	1	5	37	32	28	26	24	20	19	25
19	20	6	9										

क्रियात्मक गतिविधि

- अपनी पुरानी अंक सारणियों से, अपनी पूर्व कक्षा में अर्द्धवार्षिक तथा वार्षिक परीक्षाओं में प्राप्त गणित के प्राप्तांकों को पता कीजिए। इन्हें वर्ष के क्रम में व्यवस्थित कीजिए। अब यह जाँच कीजिए कि क्या उक्त विषय में आप द्वारा प्राप्त किए अंक चर हैं या नहीं। इसके साथ यह भी देखिए कि क्या बाद के वर्षों में गणित में आपकी स्थिति में सुधार हुआ है?

© NCERT not to be republished