

इकाई की रूपरेखा

- 19.1 प्रस्तावना
- 19.2 सहसंबंध
- 19.3 अवर्गीकृत आंकड़ों के सहसंबंध की गणना की विधियां धियां
- 19.4 वर्गीकृत आंकड़ों के सहसंबंध गुणांक की परिकलित करने की विधियां
- 19.5 समाश्रयण
- 19.6 निष्कर्ष
- 19.7 कुछ उपयोगी पुस्तकें

अध्ययन के उद्देश्य

यह आशा की जाती है कि इकाई 19 को पढ़ लेने के बाद आपके लिए संभव होगा:

- दो या अधिक चरों के बीच के सह विचरण विश्लेषण के महत्व को समझना;
- विभिन्न प्रकार के सहसंबंधों की व्याख्या करना;
- अवर्गीकृत और वर्गीकृत दोनों ही आंकड़ों के सहसंबंध की परिकलन विधियां बताना; तथा
- समाश्रयण विश्लेषण की विधि को समझना। यह एक या अधिक चर ज्ञात होने पर एक चर के मानों का आकलन करने में सहायक होती है।

19.1 प्रस्तावना

इकाई 18 के अंतिम भाग में हमने साध्यों के बीच के संबंधों का उल्लेख किया है। आइए अब हम सहसंबंध और समाश्रयण के विषय पर चर्चा करें।

इकाई 19 सहसंबंध के बारे में है जो कि दो या अधिक चरों के बीच के सह-विचरण का एक विश्लेषण है। इस इकाई का अध्ययन करने पर आपको यह स्पष्ट होगा कि सहसंबंध का सांख्यिकीय साधन दो चरों के बीच के मात्रात्मक संबंध को मापने और व्यक्त करने में सहायक होता है। इकाई 19 में इस सांख्यिकीय साधन को लागू करने की विधियों की व्याख्या की गई है। इसमें सामाजिक विज्ञान में सहसंबंध गुणांक, निर्धारण गुणांक और समाश्रयण विश्लेषण का महत्व दर्शाया गया है। इसमें समाश्रयण विश्लेषण की व्याख्या भी की गई है जो कि एक या अधिक चरों को ज्ञात होने पर एक चर के मानों का आकलन करने की एक विधि है। इस बात की ओर ध्यान दिए बिना कि सहसंबंध का सांख्यिकीय साधन कठिन और जटिल होता है। इस इकाई में इस साधन को लागू करने के बारे में बताया गया है।

19.2 सहसंबंध (Correlation)

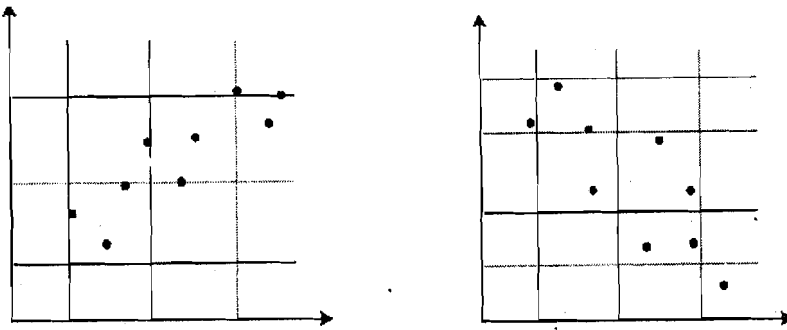
सहसंबंध दो या अधिक चरों के बीच के सह-विचरण का एक विश्लेषण है। जब दो चरों के बीच का संबंध मात्रात्मक होता है तो इस संबंध को मापने का और इसे एक संक्षिप्त सूत्र में व्यक्त करने के सांख्यिकीय साधन को सहसंबंध कहा जाता है। यदि एक चर में परिवर्तन होने पर दूसरे चर में संगत परिवर्तन होता है, तो ये दो चर सह-संबंधित होते हैं। आइए हम सहसंबंध के प्रकार पर चर्चा करें।

सहसंबंध के प्रकार (Types of correlation)

सहसंबंध के प्रकार पर विचार करते समय हमने दो प्रकार के सहसंबंधों पर चर्चा की है:

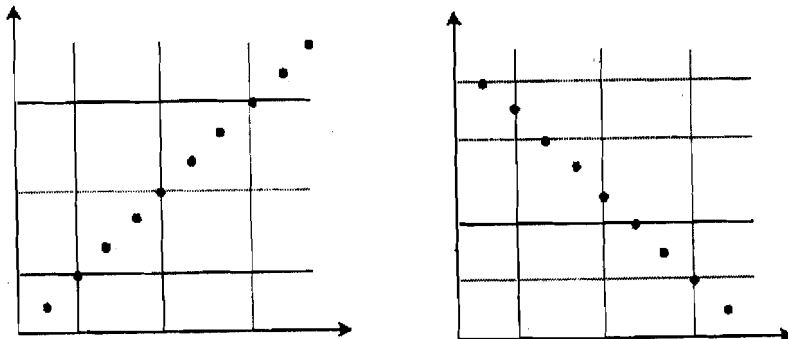
- क) धनात्मक (positive) और ऋणात्मक (negative) सहसंबंध
- ख) रैखिक (linear) और अरैखिक (non-linear) सहसंबंध
- क) धनात्मक और ऋणात्मक सहसंबंध

यदि दो चरों के मान एक ही दिशा में विचलित होते हों अर्थात् यदि एक चर के मान में वृद्धि होने पर दूसरे चर के मान में औसतन संगत वृद्धि होती हो या एक चर के मान में कमी होने पर दूसरे चर के मान में कमी होती हो, तो सहसंबंध को धनात्मक या अनुलोम (direct) सहसंबंध कहा जाता है। धनात्मक सहसंबंध के कुछ उदाहरण हैं: (i) ऊँचाई और भार (ii) भूमि का स्वामित्व और घरेलू आय। इसके विपरीत, यदि चर विपरीत दिशाओं में विचलित होते हों अर्थात् यदि एक चर के मान में वृद्धि (कमी) होने पर दूसरे चर के मान में कमी (वृद्धि) होती हो, तो इस स्थिति में सहसंबंध को ऋणात्मक या प्रतिलोम (indirect) सहसंबंध कहा जाता है। ऋणात्मक सहसंबंध के कुछ उदाहरण हैं: (i) भौतिक संपत्ति और गरीबी का स्तर, (ii) शरीर में ताकत और आयु। चित्र 19.1 में धनात्मक और ऋणात्मक प्रकार के सहसंबंध दिखाए गए हैं।



चित्र 19.1: (क) धनात्मक सहसंबंध और (ख) ऋणात्मक सहसंबंध

सहसंबंध के मान -1 से $+1$ के बीच होता है। जब $r = +1$, तो इसका अर्थ यह है कि चरों के बीच परिपूर्ण धनात्मक सहसंबंध है और जब $r = -1$, तो इसका अर्थ यह है कि चरों के बीच परिपूर्ण ऋणात्मक सहसंबंध है। जब $r = 0$, तो इसका अर्थ यह है कि दो चरों के बीच कोई सहसंबंध नहीं है (देखिए चित्र 19.2)



चित्र 19.2: (क) परिपूर्ण धनात्मक सहसंबंध ($r = +1$) और (ख) परिपूर्ण ऋणात्मक सहसंबंध ($r = -1$)

ख) रैखिक और अरैखिक सहसंबंध: (Linear and Non-linear correlation)

दो चरों के बीच के सहसंबंध को तब रैखिक कहा जाता है जबकि एक चर में इकाई परिवर्तन होने पर मानों के पूरे परिसर (range) के दूसरे चर में एक नियत परिवर्तन होता हो। चित्र 19.3 में दिए गए इन आंकड़ों को लीजिए।

x	1	2	3	4	5	6
y	3	5	7	9	11	13

चित्र 19.3: मानों के पूरे परिसर में दिखाया गया नियत परिवर्तन

इस स्थिति में चित्र 19.3 में दिखाए गए आंकड़ों के संबंध $y = 1 + 2x$ से निरूपित किया जा सकता है। सामान्यतया दो चरों को रैखिकतः संबंधित तब कहा जाता है जबकि इन चरों के बीच $y = a + bx$ के रूप का संबंध होता है।

इसके विपरीत, दो चरों के बीच के संबंध को अरैखिक या वक्ररेखी (curvilinear) तब कहा जाता है जबकि एक चर में इकाई परिवर्तन होने पर दूसरे चर में नियत या उच्चावचन (fluctuating) दर से परिवर्तन न होता हो। अरैखिक सहसंबंध का उदाहरण चित्र 19.4 में निम्नलिखित आंकड़ा समुच्चय (data set) से दिया गया है।

x	1	2	3	4	5	6
y	5	8	14	15	18	22

चित्र 19.4: अरैखिक सहसंबंध

क्योंकि चित्र 19.4 के उदाहरण में x के मान में इकाई परिवर्तन होने पर y के मान में उच्चावचन (नियत नहीं) परिवर्तन होता है, अतः यह एक अरैखिक सहसंबंध को निरूपित करता है।

अब आपको यह जानने की इच्छा होगी कि सहसंबंध का अध्ययन किस प्रकार किया जाता है। आइए हम सहसंबंध का अध्ययन करने की विधियों पर संक्षेप में चर्चा करें। परन्तु, सहसंबंध का अध्ययन करने की विधियों पर चर्चा करने से पहले आइए हम साँचें और करें 19.1 को पूरा करें।

अनुचिंतन और किया 19.1

अपनी परिकल्पना (hypothesis) से संबंधित धनात्मक और ऋणात्मक सहसंबंधों का चित्र बनाइए। इसके बाद परिपूर्ण धनात्मक और परिपूर्ण ऋणात्मक संबंधों का एक अन्य चित्र बनाइए। इसके अतिरिक्त, मानों के पूरे परिसर में प्रतिबिंबित नियत परिवर्तन और अरैखिक सहसंबंध के दो और चित्र बनाइए। चित्रों को बनाने के लिए ऊपर दिए गए 18.1 से 18.4 तक के चित्रों की सहायता आप लें।

सहसंबंध का अध्ययन करने की विधियाँ

दो चरों के बीच कोई सहसंबंध है या नहीं, इसे ज्ञात करने की विभिन्न विधियाँ हैं (i) प्रकीर्ण आरेख (scatter diagram);(ii) ग्राफीय विधि (graphic diagram);(iii) कार्ल पियर्सन का सहसंबंध गुणांक (coefficient of correlation) ; (iv) कोटि विधि (rank method); (v) संगामी विचलन विधि (concurrnt deviation method); और (vi) न्यूनतम वर्ग विधि (method of least squares)।

इनमें से पहली दो विधियाँ आरेखों और ग्राफों के ज्ञान पर आधारित हैं और शेष विधियाँ गणितीय साधनों पर आधारित हैं। प्रयुक्त किए गए अनेक गणितीय साधनों में अति

लोकप्रिय साधन हैं कार्ल पियर्सन का सहसंबंध गुणांक (r)। अतः हमने इस विधि पर विशेष ध्यान दिया है अवर्गीकृत आंकड़ों (ungrouped data) और वर्गीकृत आंकड़ों से सहसंबंध का परिकलन करने की क्रियाविधि अलग-अलग है।

19.3 अवर्गीकृत आंकड़ों के सहसंबंध की गणना की विधियाँ

अवर्गीकृत आंकड़ों से सहसंबंध गुणांक का परिकलन करने की विभिन्न विधियाँ हैं:

- i) वास्तविक माध्य के प्रयोग से
- ii) कल्पित माध्य के प्रयोग से
- iii) अनुलोम विधि (direct method)

इन सभी विधियों के प्रयोग को नीचे दिए गए उदाहरण की सहायता से दर्शाया गया है।

उदाहरण: नीचे चित्र 19.5 में दिए गए आंकड़ों की सहायता से पतियों और पत्नियों की विवाह के समय की आयु के बीच (कार्ल पियर्सन) सह-संबंध गुणांक ज्ञात कीजिए।

विवाह के समय की आयु	स्थिति1	स्थिति2	स्थिति3	स्थिति4	स्थिति5	स्थिति6	स्थिति7	स्थिति8	स्थिति9	स्थिति10
पति	28	25	24	29	31	22	21	25	26	28
पत्नी	22	23	21	25	26	20	19	21	21	24

चित्र 19.5: पतियों और पत्नियों की विवाह के समय की आयु के बीच का सहसंबंध गुणांक

वास्तविक माध्य के प्रयोग से सहसंबंध गुणांक का परिकलन करने की विधि

पहले वास्तविक माध्य का प्रयोग करके सहसंबंध गुणांक का परिकलन करने की विधि का अध्ययन कर लेने के बाद आपके लिए आपने आप ही सहसंबंध गुणांक का परिकलन करना संभव हो जाएगा।

r का परिकलन करने के लिए इस सूत्र को लागू किया जाता है

$$r = \frac{\sum xy}{N * \delta_x * \delta_y}$$

जहाँ $x = (X - M_x)$, जहाँ M_x , X मानों की श्रेणी का माध्य है

$y = (Y - M_y)$ जहाँ M_y , Y मानों की श्रेणी माध्य है

δ_x = श्रेणी X का मानक विचलन

δ_y = श्रेणी Y का मानक विचलन

N = प्रेक्षण-युग्मों की संख्या

सूत्र को इस प्रकार भी व्यक्त किया जा सकता है:

$$r = \frac{\sum xy}{\sqrt{[\delta_x^2 * \delta_y^2]}}$$

निम्नलिखित चरणों में सहसंबंध गुणांक का परिकलन को दर्शाया गया है।

- i) X के माध्य से X श्रेणी के विचलन लीजिए और उन्हें x से प्रकट कीजिए।
- ii) इन विचलनों का वर्णन कीजिए और कुल योग अर्थात् δ_x^2 प्राप्त कीजिए:
- iii) Y के माध्य से Y श्रेणी के विचलन लीजिए और इन्हें y से प्रकट कीजिए।
- iv) इन विचलनों का वर्ग कीजिए और कुल योग अर्थात् δ_y^2 प्राप्त कीजिए।

v) x और y के विचलनों को गुणा कीजिए और कुल योग $\sum xy$ प्राप्त कीजिए; और

vi) ऊपर के सूत्र में $\sum x^2$, $\sum y^2$ और $\sum xy$ के मान प्रतिस्थापित कीजिए।

वास्तविक माध्य का प्रयोग करके सहसंबंध गुणांक का परिकलन

इस विधि का अध्ययन कर लेने के बाद आइए अब हम परिकलन करें जैसा कि चित्र 19.6 में दिखाया गया है।

X	$x = X - M_x$	X^2	Y	$y = Y - M_y$	y^2	xy
28	2.1	04.41	22	-0.2	00.04	-00.42
25	-0.9	00.81	23	0.8	00.64	-00.72
24	-1.9	03.61	21	-1.2	01.44	02.28
29	3.1	09.61	25	2.8	07.84	08.68
31	5.1	26.01	26	3.8	14.44	19.38
22	-3.9	15.21	20	-2.2	04.84	08.58
21	-4.9	24.01	19	-3.2	10.24	15.68
25	-0.9	00.81	21	-1.2	01.44	01.08
26	0.1	00.01	21	-1.2	01.44	-00.12
28	2.1	04.41	24	1.8	03.24	03.78
259	0	88.90	222	0	45.60	58.20

चित्र 19.6: वास्तविक माध्य के प्रयोग से सहसंबंध गुणांक का परिकलन

$$r = \frac{\sum xy}{\sqrt{[\sum x^2 * \sum y^2]}}$$

$$M_x = 259 / 10 = 25.9 \quad M_y = 222 / 10 = 22.2 \quad (\sum x^2) = 88.9 \quad (\sum y^2) = 45.6 \quad \sum xy = 58.2$$

$$r = 58.2 / \sqrt{[88.9 * 45.6]} = 0.914$$

कल्पित माध्य के प्रयोग से सहसंबंध गुणांक का परिकलन करने की विधि

ऊपर बतायी गई विधि और इस विधि में अंतर केवल यह है कि ऊपर बताई गई विधि में विचलन वास्तविक माध्य से लिए गए हैं जबकि इस विधि में विचलन कल्पित माध्य से लिए गए हैं (अर्थात् x की श्रेणी और y की श्रेणी को देखकर x और y के माध्य कल्पित कर लीजिए और फिर वही विधि लागू कीजिए जो कि ऊपर बतायी गई विधि में लागू की गई है।

कल्पित माध्य के प्रयोग से सहसंबंध गुणांक का परिकलन

अब हम चित्र 19.7 के अनुसार परिकलित करें।

X	$D_x = X - A_x$	d_x^2	Y	$d_y = Y - A_y$	d_y^2	$d_x * d_y$
28	3	9	22	0	0	0
25	0	0	23	1	1	0
24	-1	1	21	-1	1	1
29	4	16	25	3	9	12
31	6	36	26	4	16	24
22	-3	9	20	-2	4	6
21	-4	16	19	-3	9	12
25	0	0	21	-1	1	0
26	1	1	21	-1	1	-1
28	3	9	24	2	4	6
259	9	97	222	2	46	60

चित्र 19.7: कल्पित माध्य के प्रयोग से सहसंबंध गुणांक का परिकलन

$$r = \frac{N \sum d_x * d_y - (\sum d_x * \sum d_y)}{\sqrt{\{N \sum d_x^2 - (\sum d_x)^2\} * \{N \sum d_y^2 - (\sum d_y)^2\}}}$$

$$r = \frac{10 * 60 - (9 * 2)}{\sqrt{\{10 * 97 - (9)^2\} * \{10 * 46 - (2)^2\}}}$$

$$r = \frac{582}{636.697}$$

$$r = 0.914$$

सहसंबंध गुणांक का परिकलन करने की प्रत्यक्ष विधि

वास्तविक अथवा कल्पित माध्य से विचलन लिए बिना ही x और y के वास्तविक मानों को लेकर भी सहसंबंध गुणांक परिकलित किया जा सकता है। निम्नलिखित सूत्र से इसका परिकलन किया जाता है:

$$r = (N * \sum XY - \sum X * \sum Y) / \sqrt{[N * \sum X^2 - (\sum X)^2] * [N * \sum Y^2 - (\sum Y)^2]}$$

प्रत्यक्ष विधि का प्रयोग करने से वही परिणाम प्राप्त होता है जो कि वास्तविक या कल्पित माध्य से लिए गए विचलनों से प्राप्त होता है। इस उदाहरण को चित्र 19.8 में दर्शाया गया है।

X	Y	X ²	Y ²	XY
28	22	784	484	616
25	23	625	529	575
24	21	576	441	504
29	25	841	625	725
31	26	961	676	806
22	20	484	400	440
21	19	441	361	399
25	21	625	441	525
26	21	676	441	546
28	24	784	576	672
259	222	6797	4974	5808

चित्र 19.8: प्रत्यक्ष विधि के प्रयोग से सहसंबंध गुणांक का परिकलन

आइए अब हम सोचें और करें 19.2 पूरा करें और तब भाग 19.4 में वर्गीकृत आंकड़ों का सहसंबंध गुणांक का परिकलन करने की विधियों का अध्ययन करें।

सोचें और करें 19.2

निम्नलिखित दो परिकलनों में से एक परिकलन लीजिए और उसे अपनी परिकल्पना में लागू कीजिए। परिकलनों में हुई त्रुटियों को लेकर आपको चिन्ता करने की कोई आवश्यकता नहीं है, क्योंकि इस समय आपका लक्ष्य क्रियाविधि को समझना है। यह आपके रिपोर्ट का एक अंश नहीं है।

- कल्पित माध्य के प्रयोग से सहसंबंध गुणांक का परिकलन
- प्रत्यक्ष विधि से सहसंबंध गुणांक का परिकलन

19.4 वर्गीकृत आंकड़ों के सहसंबंध गुणांक की परिकलित करने की विधियाँ

जब प्रेक्षणों की संख्या बृहत् होती है, तो आंकड़ों को द्विधा बारंबारता-बंटन (two-way frequency distribution), जिसे सहसंबंध सारणी कहा जाता है, के रूप में प्रस्तुत किया जाता है। y श्रेणी के वर्ग-अंतरालों को स्तंभ शीर्षक के रूप में लिखा जाता है और श्रेणी को पंक्ति शीर्षक के रूप में लिखा जाता है। दो चरों के बारंबारता-बंटन को उनकी कोष्ठिकाओं में लिखा जाता है। सहसंबंध गुणांक परिकलित करने का सूत्र यह है:

$$r = \frac{\sum f \cdot d_x \cdot d_y - (\sum f_x \cdot d_x \cdot \sum f_y \cdot d_y) / N}{\sqrt{\{\sum f_x \cdot d_x^2 - (\sum f_x \cdot d_x)^2 / N\} \cdot \sqrt{\{\sum f_y \cdot d_y^2 - (\sum f_y \cdot d_y)^2 / N\}}}$$

चरण:

- i) चर x के सोपानी विचलन लीजिए और इन विचलनों को d_x से प्रकट कीजिए।
- ii) चर y के सोपानी विचलन लीजिए और इन विचलनों को d_y से प्रकट कीजिए।
- iii) $d_x \cdot d_y$ और प्रत्येक कोष्ठिका (cell) की बारंबारताओं को गुणा कीजिए और प्राप्त अंक को कोष्ठिका की दायीं ओर के ऊपरी कोनों में लिख दीजिए।
- iv) सभी मानों को एक साथ जोड़कर $\sum f \cdot d_x \cdot d_y$ प्राप्त कीजिए।
- v) चर x की सभी बारंबारताओं को x के विचलनों से गुणा कीजिए और कुल योग $\sum f_x \cdot d_x$ प्राप्त कीजिए।
- vi) चर x के विचलनों का वर्णन कीजिए और उनकी बारंबारताओं से गुणा करके $\sum f_x \cdot d_x^2$ प्राप्त कीजिए।
- vii) चर y की सभी बारंबारताओं को y के विचलनों से गुणा कीजिए और कुल योग $\sum f_y \cdot d_y$ प्राप्त कीजिए।
- viii) चर y के विचलनों का वर्णन कीजिए और उनकी बारंबारताओं से गुणा करके $\sum f_y \cdot d_y^2$ प्राप्त कीजिए।
- ix) ऊपर के सूत्र में $\sum f_y \cdot d_y^2$, $\sum f_y \cdot d_y$, $\sum f_x \cdot d_x^2$, $\sum f_x \cdot d_x$, $\sum f \cdot d_x \cdot d_y$ के मान प्रतिस्थापित करके r का मान प्राप्त कीजिए।

आइए अब हम चित्र 19.9 में दिए गए आंकड़ों से कार्ल पियर्सन का सहसंबंध गुणांक परिकलित करने से संबंधित एक उदाहरण लें।

विलास की वस्तुओं पर खर्च	आय (हजार रुपयों में)				
	20-25	25-30	30-35	35-40	40-45
0-4	28	12	05		
4-8	41	22	09	03	
8-12	05	33	28	14	16
12-16	18	22	29	37	
16-20	03	09	12		

चित्र 19.9: विलास की वस्तुओं पर किए गए खर्च से संबंधित सहसंबंध गुणांक

हमारे लिए प्रत्यक्ष विधि से वर्गीकृत आंकड़ों का सहसंबंध गुणांक परिकलित करना संभव है, जैसा कि चित्र 19.10 में दिखाया गया है। (देखिए चित्र 19.10)

सहसंबंध और समाश्रयण

विलास की वस्तुओं पर खर्च	आय (हजार रुपयों में)								
	20-25	25-30	30-35	35-40	40-45	f_y	DY	$f_y \cdot d_y$	$f_y \cdot d_y \cdot d_y$
0-4	28	12	5			45	-2	-90	180
4-8	41	22	9	3		75	-1	-75	75
8-12	9	33	28	14	16	100	0	0	0
12-16		18	22	29	37	106	1	106	106
16-20			3	9	12	24	2	48	96
F_x	78	85	67	55	65	350		-11	457
d_x	-2	-1	0	1	2				
$f_x \cdot d_y$	-156	-85	0	55	130	-56			
$f_x \cdot d_x \cdot d_x$	312	85	0	55	260	712			

चित्र 19.10: वर्गीकृत आंकड़ों के सहसंबंध गुणांक का परिकलन

अब हमारे प्रत्यक्ष विधि से $f_x \cdot d_x \cdot d_y$ का परिकलन करना संभव है जैसा कि चित्र 19.11 में दिया गया है।

विलास की वस्तुओं पर खर्च	आय (हजार रुपयों में)					$f_x \cdot d_x \cdot d_y$
	20-25	25-30	30-35	35-40	40-45	
0-4	112	24	0	0	0	136
4-8	82	22	0	-3	0	101
8-12	0	0	0	0	0	0
12-16	0	-18	0	29	74	85
16-20	0	0	0	18	48	66
$f_x \cdot d_x \cdot d_y$	194	28	0	44	122	388

चित्र 19.11: वर्गीकृत आंकड़ों के सहसंबंध गुणांक का परिकलन

$$N = 350 \quad \Sigma f_x \cdot d_x \cdot d_y = 388 \quad \Sigma f_x \cdot d_x = -56$$

$$\Sigma f_y \cdot d_y = -11 \quad \Sigma f_x \cdot d_x^2 = 712 \quad \Sigma f_y \cdot d_y^2 = 457$$

$$r = \frac{\Sigma f_x \cdot d_x \cdot d_y - (\Sigma f_x \cdot d_x \cdot \Sigma f_y \cdot d_y) / N}{\sqrt{\{\Sigma f_x \cdot d_x^2 - (\Sigma f_x \cdot d_x)^2 / N\} \cdot \sqrt{\{\Sigma f_y \cdot d_y^2 - (\Sigma f_y \cdot d_y)^2 / N\}}}$$

$$r = \frac{388 - (-56 \cdot -11) / 350}{\sqrt{\{712 - (-56)^2 / 350\} \cdot \sqrt{\{457 - (-11)^2 / 350\}}}$$

$$r = 386.24 / (26.515 \cdot 21.369) = .682$$

अधिकांश चरों में किसी न किसी प्रकार का संबंध अवश्य होता है। सहसंबंध की सहायता से दो या अधिक चरों के बीच के संबंध की कोटि मापी जा सकती है। हमें मालूम है कि

सहसंबंध से कारण और प्रभाव संबंध के बारे में कुछ भी पता नहीं चलता। यहां तक कि उच्च कोटि के संबंध भी यह नहीं बताता कि कारण और प्रभाव संबंध का अस्तित्व है या नहीं। जबकि विलोमतः कारण और प्रभाव संबंध (या प्रकार्य संबंध) से सदैव ही सहसंबंध का व्यंजक प्राप्त हो जाता है। अब समाश्रयण विश्लेषण पर चर्चा की जाएगी।

19.5 समाश्रयण (Regression)

समाश्रयण[⊙] विश्लेषण (regression analysis) एक या अधिक चरों के ज्ञान से एक चर के मानों का आकलन करने की विधि है। अनुसंधानकर्ता जिस चर का आकलन करने का प्रयास करता है उसे आश्रित चर (dependent variable) कहा जाता है और इसे y से प्रकट किया जाता है जबकि प्रागुक्ति (prediction) के लिए जिस चर का प्रयोग किया जाता है उसे स्वतंत्र चर (independent variable) कहा जाता है और इसे x से प्रकट किया जाता है। समाश्रयण समीकरण में एक या अधिक स्वतंत्र चर हो सकते हैं, परन्तु आश्रित चर केवल ही होता है। समाश्रयण समीकरण में एक चर है या अधिक चर हैं इसके अनुसार समाश्रयण को सरल या बहुल कहा जाता है। जब आश्रित और स्वतंत्र चर के बीच का संबंध रैखिक होता है तो इन समीकरणों में आगे एक शब्द 'रैखिक' जोड़ दिया जाता है। इस तरह, एक सरल रैखिक समाश्रयण समीकरण को इस प्रकार व्यक्त किया जाता है:

$$Y = a + bX$$

जहां Y आश्रित चर है

x स्वतंत्र चर है

' a ' समाश्रयण अचर हैं

' b ' समाश्रयण गुणांक हैं। यह X में हुए परिवर्तन संगत Y में हुए परिवर्तन को मापता है।

इसी प्रकार बहुरैखिक समाश्रयण समीकरण को इस प्रकार व्यक्त किया जाता है

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n$$

जहां Y आश्रित चर हैं

X_1, X_2, \dots, X_n स्वतंत्र चर हैं

' a ' समाश्रयण अचर हैं

b_1, b_2, \dots, b_n समाश्रयण गुणांक हैं।

सहसंबंध गुणांक का परिकलन करने की भांति समाश्रयण समीकरण का परिकलन करने की विभिन्न विधियाँ होती है।

1) X और Y के वास्तविक माध्य मानों से

2) X और Y के कल्पित माध्य मानों से

वास्तविक माध्य के प्रयोग से समाश्रयण समीकरण का परिकलन

निम्नलिखित सूत्र को लागू करके (X पर Y के) समाश्रयण समीकरण का परिकलन किया जा सकता है।

$$Y - M_y = b_{yx} * (X - M_x) \text{ या}$$

$$Y - M_y = r(\delta_y / \delta_x) * (X - M_x)$$

$$\text{As, } b_{yx} = r(\delta_y / \delta_x) = (?xy / ?x^2)$$

इसलिए निम्नलिखित सूत्र को लागू करके समाश्रयण समीकरण परिकलित किया जा सकता है।

$$Y - M_y = (\Sigma xy / \Sigma x^2) * (X - M_x)$$

जहाँ Y और X क्रमशः आश्रित चर और स्वतंत्र चर हैं।

M_y और M_x और चर Y और X और के क्रमशः माध्य है, और

$$y = Y - M_y \text{ और } x = X - M_x$$

नीचे के उदाहरण में समाश्रयण समीकरण का परिकलन दर्शाया गया है।

उदाहरण: पतियों के विवाह के समय की आयु को स्वतंत्र चर मानकर और पत्नियों की आयु को आश्रित चर मानकर निम्नलिखित आंकड़ों से समाश्रयण समीकरण परिकलित कीजिए (देखिए चित्र 19.11)

विवाह के समय की आयु	स्थिति1	स्थिति2	स्थिति3	स्थिति4	स्थिति5	स्थिति6	स्थिति7	स्थिति8	स्थिति9	स्थिति10
पति	28	25	24	29	31	22	21	25	26	28
पत्नी	22	23	21	25	26	20	19	21	21	24

वास्तविक माध्य के प्रयोग से समाश्रयण समीकरण का परिकलन (देखिए चित्र 19.12)

पत्नी की आयु Y	$y = Y - M_y$	y^2	पति की आयु X	$X = X - M_x$	x^2	xy
22	-0.2	00.04	28	2.1	4.41	-0.42
23	0.8	00.64	25	-0.9	0.81	-0.72
21	-1.2	01.44	24	-1.9	3.61	02.28
25	2.8	07.84	29	3.1	9.61	08.68
26	3.8	14.44	31	5.1	26.01	19.38
20	-2.2	04.84	22	-3.9	15.21	8.58
19	-3.2	10.24	21	-4.9	24.01	15.68
21	-1.2	01.44	25	-0.9	0.81	01.08
21	-1.2	01.44	26	0.1	0.01	-0.12
24	1.8	03.24	28	2.1	4.41	03.78
222	0	45.60	259	0	88.9	58.20

चित्र 19.12: वास्तविक माध्य के प्रयोग से समाश्रयण समीकरण

$$M_y = 222 / 10 = 22.2 \quad M_x = 259 / 10 = 25.9$$

$$Y - M_y = (\Sigma xy / \Sigma x^2) * (X - M_x)$$

$$Y - 22.2 = (58.2 / 88.9) * (X - 25.2)$$

$$Y - 22.2 = 0.655 * (X - 25.2)$$

$$Y - 22.2 = 0.655X - 16.96$$

$$Y = 5.24 + 0.655X$$

कल्पित मान के प्रयोग से समाश्रयण समीकरण का परिकलन (देखिए चित्र 19.13)

कल्पित माध्य लेकर निम्नलिखित सूत्र से (X पर Y का) समाश्रयण समीकरण परिकलित किया जा सकता है।

$$Y - M_Y = b_{yx} * (X - M_x)$$

जहाँ

$$b_{yx} = [\Sigma d_x * d_y - (\Sigma d_x * \Sigma d_y) / N] / [\Sigma d_x^2 - (\Sigma d_x)^2 / N]$$

Y और X क्रमशः आश्रित चर और स्वतंत्र चर हैं।

M_y और M_x , चरों Y और X के क्रमशः माध्य हैं

$$d_y = Y - AM_y \text{ और } d_x = X - AM_x$$

AM_y और AM_x चरों Y और X के क्रमशः कल्पित माध्य हैं; और

कल्पित माध्य के प्रयोग से समाश्रयण समीकरण का परिकलन

Age of Wives	$d_y = Y - AM_y$	d_y^2	Age of Husbands	$d_x = X - AM_x$	d_x^2	$d_x * d_y$
22	0	0	28	3	9	0
23	1	1	25	0	0	0
21	-1	1	24	-1	1	1
25	3	9	29	4	16	12
26	4	16	31	6	36	24
20	-2	4	22	-3	9	6
19	-3	9	21	-4	16	12
21	-1	1	25	0	0	0
21	-1	1	26	1	1	-1
24	2	4	28	3	9	6
222	2	46	259	9	97	60

चित्र 19.13: कल्पित माध्य के प्रयोग से समाश्रयण समीकरण का परिकलन

$$M_Y = 222 / 10 = 22.2 \quad M_x = 259 / 10 = 25.9$$

$$b_{yx} = [\Sigma d_x * d_y - (\Sigma d_x * \Sigma d_y) / N] / [\Sigma d_x^2 - (\Sigma d_x)^2 / N]$$

$$b_{yx} = [60 - (9*2) / 10] / [97 - 9*9 / 10]$$

$$b_{yx} = 58.2 / 88.9 = 0.655$$

$$Y - M_Y = b_{yx} * (X - M_x)$$

$$Y - 22.2 = 0.655 * (X - 25.2)$$

$$Y - 22.2 = 0.655X - 16.96$$

$$Y = 5.24 + 0.655X$$

आकलन की मानक त्रुटि: समाश्रयण समीकरण की सहायता से तब तक परिपूर्ण प्रागुक्ति नहीं की जा सकती जब तक कि सहसंबंध का मान -1 या +1 न हो। इस तरह, अनुसंधानकर्ता की रुचि एक समाश्रयण समीकरण के आकलन की परिशुद्धता ज्ञात करने में होती है। आकलन की मानक त्रुटि आकलन के एक आधार के रूप में समाश्रयण समीकरण का प्रयोग करने से संबंधित त्रुटि मापती है। निम्नलिखित समीकरण को लागू करके इसे परिकलित किया जा सकता है।

$$SEE_{y,x} = \sqrt{\frac{\sum(Y - Y_c)^2}{N - 2}}$$

जहाँ $SEE_{y,x}$ आकलन की मानक त्रुटि है।

Y आश्रित चर है

Y_c , Y का प्रागुक्त मान है

N प्रेक्षणों की संख्या है

निम्नलिखित सूत्र से भी इसे परिकलित किया जा सकता है।

$$SEE_{y,x} = \sqrt{\frac{(\sum Y^2 - a^2 Y - b \sum XY) / N - 2}{N}}$$

जहाँ $SEE_{y,x}$ आकलन की मानक त्रुटि है।

Y आश्रित चर है

X स्वतंत्र चर है 'a' समाश्रयण अचर है

'b' समाश्रयण गुणांक है

N प्रेक्षणों की संख्या है

निर्धारण-गुणांक (coefficient of determination): निर्धारण-गुणांक (r^2) सहसंबंध गुणांक (r) का वर्ग होता है और प्रायः इसका प्रयोग सहसंबंध गुणांक के मान का निर्वचन (interpretation) करने में किया जाता है। यदि r का मान 0.8 हो, तो निर्धारण-गुणांक या r^2 का मान 0.64 होगा। इसका अर्थ यह होगा कि एक (आश्रित) चर के प्रसरण (variance) के 64% की व्याख्या अन्य (स्वतंत्र) चर के पदों में की जाती है।

सोचें और करें 19.3

हमने यह समझने का प्रयास किया है कि कल्पित माध्य के प्रयोग से समाश्रयण समीकरण का परिकलन किस प्रकार किया जाता है, परन्तु हमें सफलता नहीं मिली। संभव है कि एक उदाहरण की सहायता से आप मुझे समझा सकें एक या दो उदाहरण लेकर एक अलग कागज पर आप अपनी व्याख्या लिखें। संभव है कि मैं इसे समझ सकूँ। इसे आप एमएसओ-002 के समन्वयकर्ता के पास भेजने का प्रयास अवश्य करें।

19.6 निष्कर्ष

इकाई 19 मात्रात्मक विधियों पर खंड 5 की अंतिम इकाई है। इस खंड की सभी पांच इकाइयों में इस बात पर बल दिया गया है कि सामाजिक अनुसंधान में मात्रात्मक विधियों का प्रयोग तब करना चाहिए जबकि वे आवश्यक और प्रासंगिक हों और जिनसे अच्छे परिणाम प्राप्त हो सकते हों। कभी-कभी इनका प्रयोग गुणात्मक विधियों (qualitative methods) के संयोजन के साथ किया जा सकता है। आपको जानकारी के अभाव में मात्रात्मक विधियों से बचने की आवश्यकता नहीं है, या डरना नहीं चाहिए कि मात्रात्मक सामग्री की आवश्यकता नहीं है, समझना कठिन होता है। खंड 5 की पांचों इकाइयों में जहाँ कहीं भी संभव हो सका है और आवश्यक समझा गया है, हमने उपयुक्त उदाहरण आपको उपलब्ध कराए हैं जिससे आप उन साधनों को समझ सकें जो कि आपके शोध परियोजना कार्य में उपयोगी सिद्ध हों।

19.7 कुछ उपयोगी पुस्तकें

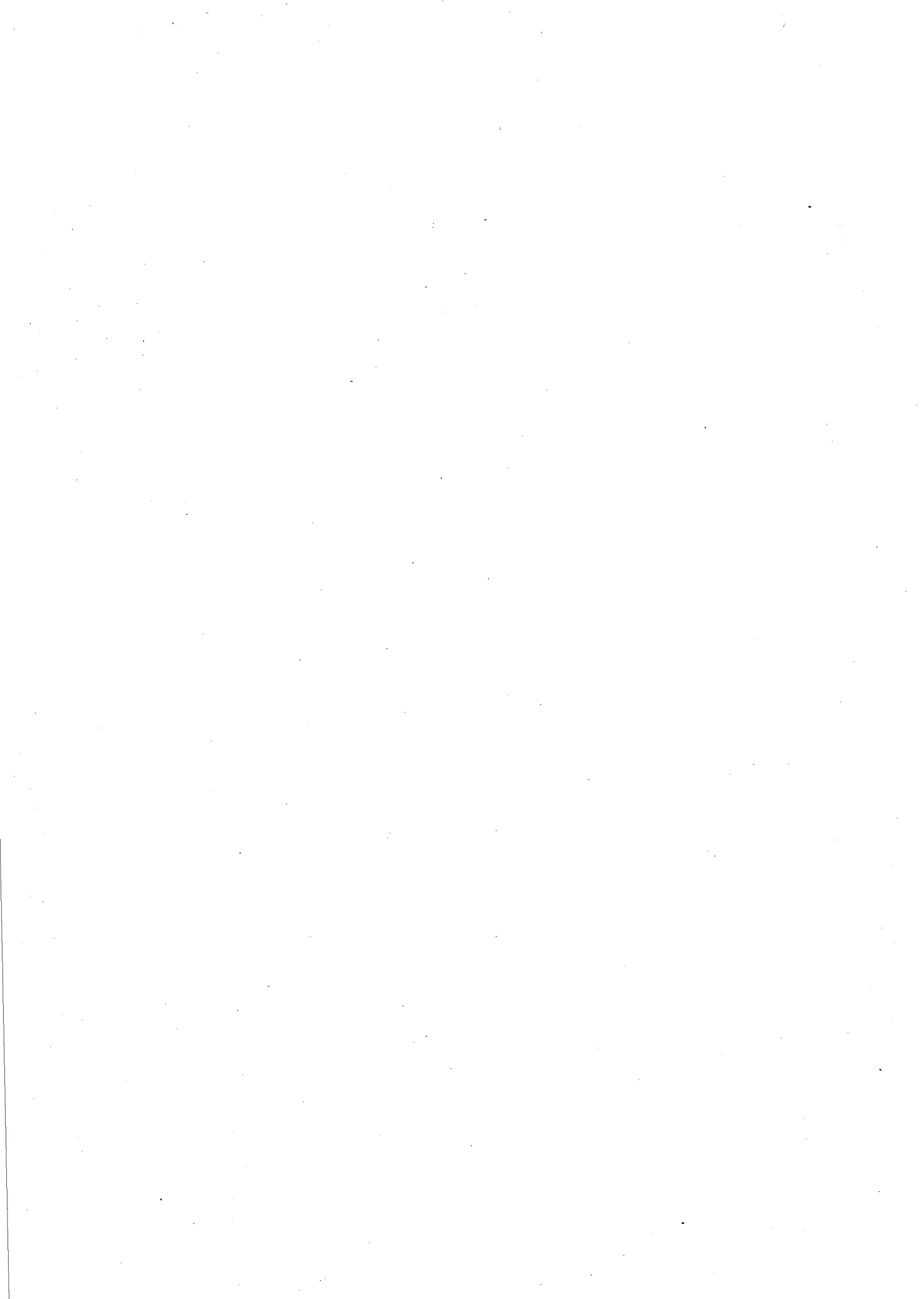
बर्नस्, रॉबर्ट बी.: 2000, *इंट्रोडक्शन टू रिसर्च मैथड्स*, सेज पब्लिकेशन: लंदन

कोहेन, लुईस और माइकेल हॉलीडे 1982, *स्टैटिस्टिक्स फॉर सोशल रिसर्च*, हार्पर और रो: लंदन

सांख्यिकी शब्दावली का हिंदी अनुवाद

अरैखिक	non-linear
अवर्गीकृत	ungrouped
आकलन	estimate
निराकरणीय परिकल्पना	null hypotheses
न्यूनतम वर्ग विधि	method of least squares
परिकल्पना	hypothesis
प्रकीर्ण आरेख	scatter diagram
प्रसरण	variance
बारंबारता	frequency
मानक त्रुटि	standard error
रैखिक	linear
वर्गीकृत	grouped
विचलन	derivation
वैकल्पिक परिकल्पना	alternative hypothesis
समाश्रयण	regression
सहसंबंध	correlation

खंड 6
सर्वेक्षण शोध



खंड 6 का परिचय

सर्वेक्षण शोध

सर्वेक्षण शोध को लोकप्रिय रूप से मतदाता सर्वेक्षण और बाजार शोध के साथ जोड़ा जाता है। दूसरी ओर, सर्वेक्षण सामाजिक शोध की पारिभाषिक विशेषता हैं और ये नृशास्त्रीय जाँच का भाग भी हैं। सामाजिक विज्ञान संबंधी शोधों में किसी भी अवस्था पर किसी न किसी प्रकार का अंकीय सर्वेक्षण किया जाना एक आम बात है। प्रायः शोधकारों को समष्टि विशेष की व्याख्या करने से पहले उसका भौतिक मानचित्र तैयार करना होता है। जो कुछ भी बताया या देखा जाता है, शोधकारों को उसकी विशिष्टता को स्थापित करना होता है और शोध के दौरान बनाए गए परस्पर संबंधों की प्रामाणिकता की जाँच करनी होती है। सर्वेक्षण पद्धतियों और अन्य मात्रात्मक तकनीकों (जो खंड 5 में बताई गई हैं) का पक्ष लेने का कारण यह है कि सर्वेक्षणों से शोधकारों को विशाल एवं जटिल समाजों से एकत्रित शोध-सामग्री का प्रबंधन करने का तरीका मिलता है। शहरी क्षेत्रों से संबंधित सामाजिक मुद्दों (उदाहरण के लिए ग्रामीण जनसंख्या का औद्योगिक शहरों की ओर पलायन का विषय) पर अध्ययन करने वाले शोधकारों द्वारा प्रव्रजन विन्यास का बड़े स्तर पर सामाजिक सर्वेक्षण किया जाता है। उन्हें प्रश्नावलियों को शोध उपकरणों के रूप में प्रयोग करना होता है तथा अपनी शोध-सामग्री को कंप्यूटरों पर संसाधित करने होते हैं। सर्वेक्षणों की उपयोगिता उनके प्रयोजन के संदर्भ में आँकी जाती है।

प्रायः घटनाओं/मतों/अन्य मुद्दों के पुंजों (sets) को गिनने तथा वर्गीकृत करने के लिए सर्वेक्षणों का प्रयोग किया जाता है। जानकारी की बृहद् मात्रा को अधिक विश्वसनीयता तथा सटीकता सर्वेक्षणों से प्राप्त मात्रात्मकता से उपलब्ध होती है। आप अपनी शोध के उद्देश्यों के संदर्भ में सर्वेक्षण की उपयोगिता को मापें। हमें यह ध्यान रखना है कि हमें मात्र एक पद्धति से सभी प्रकार की जानकारी नहीं प्राप्त हो सकती है। इसके अतिरिक्त, प्रत्यक्ष रू-ब-रू साक्षात्कार, प्रेक्षण से अनुमान करना तथा सर्वेक्षणों द्वारा शीर्षों की गणना करना जैसी विभिन्न पद्धतियाँ विभिन्न परिणाम देती हैं और प्रायः एक-दूसरे की पूरक बनकर व्याख्या एवं विश्लेषण के लिए शोध-सामग्री का एक अधिक समन्वित पुंज उपलब्ध कराती हैं। मिचेल (1967 : 21) के अनुसार, मात्रात्मक विधियाँ गुणात्मक शोध के अधीनस्थ होती हैं तथा मात्रात्मक सामग्री प्राप्त करना उस समझ को प्राप्त करने का एक साधन मात्र है जो सांख्यिकी द्वारा मिलना संभव नहीं है।

साथ-साथ, मात्रात्मक विधियों के प्रति रुझान बढ़ रहा है। जैसा कि ऐलन (1984: 259) ने बताया, क्या सर्वेक्षित किया जा सकता है और 'तथ्यों' के रूप में एकत्रित किया जा सकता है तथा प्रत्यर्थियों (respondents) द्वारा बताए गए सर्वेक्षण के तथ्यों के बीच अंतर करने की हमें ज़रूरत है। पहले प्रकार में मदों की साधारण गिनती होती है तथा दूसरे प्रकार में मदों के बारे में कथन की गिनती होती है। चूँकि प्रत्यर्थियों के मत भी सामाजिक तथ्य होते हैं, इसलिए उनपर ध्यान देना महत्वपूर्ण होता है। ऐसी शोध-सामग्री को उसके सामाजिक संदर्भ में रखकर तथा अन्य प्रकार की सामग्री के बारे में जानकर ही अधिक व्याख्या तथा विश्लेषण हेतु वैज्ञानिक रूप से विश्वसनीय तथा सटीक जानकारी पाने के लक्ष्य की शोधकार की आशा पूरी हो सकती है। अन्य प्रकार की सामग्री से हमारा तात्पर्य उन 'ठोस तथ्यों' से है जो जनगणना सामग्री तथा सरकारी सर्वेक्षणों से प्राप्त होती है।

सभी सर्वेक्षणों में, हमें शोध-सामग्री एकत्रित करने और उसकी व्याख्या के तरीके पर सावधानी से ध्यान देने की आवश्यकता होती है। अन्य शब्दों में, हमें अंकों और सांख्यिकी के बीच के अंतर तथा सांख्यिकीय और समाजशास्त्रीय अनुमान के अंतर को समझना

आवश्यक है। रोचक बात है कि अधिकांश उपयोगी सर्वेक्षण गहन क्षेत्रीय-शोध कार्य द्वारा पहले शोध-क्षेत्र को समीप से जान लेने की रणनीति का प्रयोग करते हैं। इस प्रक्रिया से उन्हें आम लोगों के साथ निकट संबंध स्थापित करने का अवसर मिल जाता है। उसके बाद ही प्रश्नावली में शामिल किए जाने वाले महत्वपूर्ण प्रश्नों को बना पाना संभव हो पाता है। प्रश्नावलियों को भरने के लिए प्रत्यर्थियों का सहयोग प्राप्त करने हेतु सर्वेक्षण को प्रचारित करना तथा प्रत्यर्थियों के साथ पूर्व-संबंध स्थापित करना आवश्यक होता है। खंड 5 की निम्नलिखित चार इकाइयाँ आपको सर्वेक्षण शोध के विभिन्न पक्षों पर आपको जानकारी देंगी जिसमें प्रश्नावलियों के माध्यम से शोध-सामग्री एकत्रित करने की विधि शामिल है, जो सर्वेक्षण शोध का एक महत्वपूर्ण अंग है।

आइए खंड 6 की प्रत्येक इकाई की विषय-वस्तु को संक्षेप में देखें :

सर्वेक्षण विधि पर इकाई 20 में सर्वेक्षण के स्वरूप तथा प्रयोजन की चर्चा की गई है। इसमें सर्वेक्षण विधि के उद्द्विकास का वर्णन किया गया है और सर्वेक्षण के सामाजिक संदर्भ को जानने की प्रासंगिकता बताई गई है। सर्वेक्षणों के लाभों तथा उसमें आने वाली समस्याओं का निरूपण इस इकाई का महत्वपूर्ण अंश है।

सर्वेक्षण प्रारूप पर इकाई 21 सर्वेक्षण शोध के प्रारंभिक चरणों के बारे में है। इसमें प्रतिचयन, प्रारूप, सर्वेक्षण शोध प्रारूप को तैयार करने के तथा सर्वेक्षण का कार्यान्वयन करने के चरणों का वर्णन किया गया है। इकाई में दिए विस्तृत निर्देशों से आपके अपने शोध कार्य में इन्हें लागू करने में मदद मिलेगी।

सर्वेक्षण उपकरण पर इकाई 22 सर्वेक्षण उपकरणों (विश्वसनीयता, वैधता आदि) के प्रारूप में आने वाले मुद्दों, प्रश्नावली बनाने, सर्वेक्षण उपकरणों के प्रकारों की चर्चा करती है।

सर्वेक्षण निष्पादन और शोध-सामग्री विश्लेषण पर इकाई 23 सर्वेक्षण शोध तथा शोध-सामग्री विश्लेषण के मुद्दों तथा समस्याओं की व्याख्या करती है। इसमें सर्वेक्षण के बाद रिपोर्ट लिखने तथा निष्कर्षों को प्रस्तुत करने के बारे में उपयोगी सामग्री दी गई है।