

Data Processing → शोध प्रक्रिया में ऑकड़ों का अव्यक्तिक महत्व है। ऑकड़ों को एकत्रित करने के पश्चात् उनका सम्पादन, संसाधन वर्गीकरण, संकेतीकरण तथा कम्प्यूटरीकरण इत्यादि सम्मिलित हैं।

ऑकड़ों के संग्रह के अंतर्गत एकत्र की गई जानकारी अलग-अलग अध्ययनों में मात्रा और स्वभाव में भिन्न होती है। उदाहरणार्थ, जब प्रश्नावली और सूची के माध्यम से सर्वेक्षण किया जाता है और ऑकड़े प्राप्त किए जाते हैं तब उत्तरों में या तो सही स्थान पर सही निदान नहीं लगाया जाता या कुछ प्रश्न अनुत्तरित छोड़े जा सकते हैं या उत्तर में इस प्रकार दिए जा सकते हैं जिनमें पुनर्निर्माण की आवश्यकता है, जो कि विश्लेषण के लिए अभिकल्पित हैं।

ऑकड़ों की जांच के लिए यह भी आवश्यक है कि यह सार्थक, उपयुक्त हो और इसमें त्रुटियों का सुधार कर लिया गया है। कभी-कभी जांचकर्ता कोई त्रुटि करता है और असंभव उत्तर लिख देता है। जैसे "एक माह में आप कितना गरम मसाले का प्रयोग करते हैं?" उत्तर लिखा जाता है - एक किलो। क्या तीन सदस्यीय परिवार एक माह में एक किलो गरम मसाले का प्रयोग कर सकता है। सही उत्तर होता 0.1 किलो। अतः इस प्रकार के उत्तरों के लिए ऑकड़ों का संसाधन आवश्यक है।

ऑकड़ों का संसाधन सही संकेतीकरण तथा कम्प्यूटर में ऑकड़े देने के लिए आवश्यक है। इससे ऑकड़े पूर्ण, त्रुटिरहित, पठनीय और संकेत दिए जाने योग्य होते हैं। सम्पादन (editing) की प्रक्रिया क्षेत्र से ही प्रारंभ हो जाती है। साक्षात्कार के उपरान्त साक्षात्कारकर्ता को त्रुटियों एवं छूटी हुई सामग्री की जांच कर लेनी चाहिए जिससे फार्म को पूरा किया जा सके। सम्पादन का कार्य कई बनाने के साथ-साथ भी पूरा हो सकता है।

ऑकड़ों के संकेतीकरण का अर्थ है उत्तरदाता द्वारा दिए गए उत्तरों को संख्यात्मक मूल्यों में बदलना या आधार सामग्री के विश्लेषण में प्रयोग किए एक-चार के विभिन्न वर्गों को संख्या के रूप में रचना। सामान्यतया संकेतीकरण प्रश्न तैयार करते समय या प्रश्नावली तथा साक्षात्कार सूची को अंतिम रूप देने से पहले तैयार किया जाता है।

ऑकड़ा विश्लेषण [Data Analysis] - Statistical techniques सांख्यिकीय प्रविधि

अनुसंधान प्रश्नों के उत्तर प्राप्त करने के लिए ऑकड़ों को इनके निहित भागों में व्यवस्थित करना (to arrange) विश्लेषण कहलता है। जैसे एक शोधकर्ता किसी घटना के प्रति सकारात्मक रुझान और प्रेजुस्ट शिझा स्तर के मध्य संबंधों को तैयार हुए एक प्राकल्पना (hypothesis)

का निर्माण करता है। वह अपने अध्ययन में कॉलेजों में उत्तरदाताओं के ऑकड़ों एकत्रित करता है। इसके पश्चात् वह ऑकड़ों को विभाजित करता है और फिर इस प्रकार इनको व्यवस्थित करता है कि शोधकर्ता को इस प्रश्न का उत्तर मिल सके - क्या उच्च शिक्षा अभिवृत्तियों (Attitude) में परिवर्तन लाती है? इस विश्लेषण के साथ ऑकड़ों की व्याख्या करना भी आवश्यक होता है। व्याख्या में परिणामों का विश्लेषण किया जाता है, कुछ अनुमान लगाए जाते हैं, तत्पश्चात् संबंधों के बारे में निष्कर्ष निकाले जाते हैं। इस प्रकार व्याख्या (Interpretation) का मतलब अर्थ निकालना और उसको समझाना है। ऑकड़ों की व्याख्या दो प्रकार से होती है:

1) अध्ययन के भीतर के संबंध और इसकी आधार सामग्री की व्याख्या करके।

2) अध्ययन के परिणामों और आधार सामग्री के भीतर निकाले गए अनुभागों की तुलना सिद्धांतों और अन्य अनुसंधान निष्कर्षों से करके।
ऑकड़ों के विश्लेषण की अवस्थाएँ (Stages in Analysis)

विश्लेषण के चार चरण निम्नवत् हैं:

1) वर्गीकरण (Categorisation) - शोध समस्या और अध्ययन के उद्देश्य के अनुसार वर्गों का निर्माण किया जाता है। ये परस्पर अनर्थ (exclusive) और गहन (exhaustive) दो प्रकार के होते हैं।

2) आवृत्ति आवंटन (Frequency Distribution) - मात्रात्मक (quantitative) ऑकड़ों का वर्गों में सारणीयन (tabulation) करना आवृत्ति आवंटन कहलाता है। यह दो प्रकार का होता है:

1) प्राथमिक आवंटन - यह वर्णनात्मक होता है और प्रत्येक वर्ग में प्रकरणों (Cases) की संख्या मात्र दर्शाता है।

2) द्वितीयक आवंटन - इसमें आवृत्तियों और प्रतिशत की तुलना की दर्शाया जाता है। यह विश्लेषण संबंधों से संबद्ध होता है। जैसे पुरुषों की आवृत्ति को स्त्रियों की आवृत्ति से तुलना करना।

मापन (Measurement) → केंद्रीय प्रवृत्तियों को मापने के लिए मापन का प्रयोग किया जाता है जिसमें माध्य (Mean), माध्यिका (Median), बहुलक (Mode) की गणना होती है। इसमें माध्य का अर्थ है मापन योग्य संख्याओं के समूह का अंकगणितीय औसत, माध्यिका का अर्थ है - किसी भी मापन समूह के मध्य बिन्दुओं का माप और बहुलक का अर्थ है - मापन संख्याओं के समूह के मापन में सबसे अधिक आवृत्ति वाला अंक। मापन सह-संबंधों के गुणांक के रूप में मापा जा सकता है।

4) व्याख्या (Interpretation) - ऑकड़ों की व्याख्या वर्णनात्मक या विश्लेषणात्मक या सैद्धांतिक दृष्टि से की जा सकती है।

Statistical Techniques of Data Analysis

सांख्यिकीय सामग्री में से कुछ ऐसे मूल्यों अथवा मापों (measurement) को ढूँढ़ा जा सकता है जो संपूर्ण श्रेणी का लगभग प्रतिनिधित्व कर सकते हैं। ये ही मूल्य अथवा माप केंद्रीय प्रवृत्तियाँ (central tendencies) कहलाते हैं। जैसे चार बच्चों के समूह में बच्चों की आयु क्रमशः 10, 12, 9 और 13 वर्ष है। वैसे अगर देखा जाय तो बच्चों की आयु में मिलावट जान पड़ती है लेकिन 11 एक ऐसा अंक है जो इस समूह के बच्चों की आयु का प्रतिनिधित्व करता है। अतः 11 वर्ष इन बच्चों की केंद्रीय अथवा माध्य प्रवृत्ति को व्यक्त करता है। अर्थात् वे मूल्य जो समूह अथवा श्रेणी की योज्यता को एक अंक में ही व्यक्त कर देता है, केंद्रीय या माध्य प्रवृत्ति (Central tendencies) के माप कहलाते हैं।

केंद्रीय प्रवृत्ति को मापने के लिए मुख्यतः तीन मापक हैं :

- 1) माध्य या मध्यमान (Mean)
- 2) माध्यिका या मध्यांक (Median)
- 3) बहुलांक (Mode)

केंद्रीय प्रवृत्ति के इन तीनों मापों में केवल समय और शुद्धता का अंतर है। मध्यमान सर्वाधिक शुद्ध मान है और इसकी गणना में समय अधिक लगता है। मध्यमान (Mean) की अपेक्षा मध्यांक (Median) कम शुद्ध मान है और इसकी गणना में भी समय कम लगता है। बहुलांक (Mode) मध्यमान और मध्यांक की अपेक्षा कम शुद्ध मान है लेकिन इसकी गणना में मध्यमान और मध्यांक की अपेक्षा समय कम लगता है।

Importance of Measures of Central Tendency

केंद्रीय प्रवृत्ति के मापों का अत्यधिक महत्व है। इनका प्रयोग आँकड़ों का विश्लेषण करने में किया जाता है। समाज विज्ञान अनुसंधानों में इसका विशेष महत्व है :

- 1) केंद्रीय प्रवृत्ति के माप समूह प्राप्कों का प्रतिनिधित्व करते हैं।
- 2) ये संपूर्ण वर्ग के गुणों को संक्षेप में प्रदर्शित करते हैं।
- 3) केंद्रीय प्रवृत्ति के मापों की सहायता से दो या अधिक समूहों की तुलना सरलता से की जा सकती है।
- 4) उच्च सांख्यिकीय विश्लेषण जैसे - प्रामाणिक विचलन (Standard deviation) तथा सह-संबंध (correlation) आदि ज्ञात करने में केंद्रीय प्रवृत्ति के मापों की आवश्यकता पड़ती है।

Mean (माध्य) - माध्य को समान्तर माध्य या अंकगणीय माध्य भी कहते हैं। अंकगणीय माध्य वस्तुतः माध्यों में सबसे सरल और उन्नत माध्य माना जाता है। माध्य (Mean) वह मूल्य है जो किसी

पदमाला या श्रेणी के समस्त पदों के मूल्य के योग को उनकी संख्या से भाग देने पर प्राप्त होता है। जैसे 11 बच्चों के प्राप्तांक इस प्रकार हैं - 64, 69, 63, 60, 65, 68, 75, 55, 50, 70 और 65। इन 11 बच्चों के अंकों का समान्तर माध्य निकालने के लिए पहले इन सभी अंकों का योग निकाला जायेगा और इस योग (addition) में 11 से भाग दिया जायेगा। अर्थात्

$$\frac{704}{11} = 64 \text{ यह } 64 \text{ इसका समान्तर माध्य है।}$$

मध्यमान ज्ञात करने के लिए निम्न सूत्र (formula) का इस्तेमाल करते हैं:

$$(M) \text{ Mean} = \frac{\sum X}{N}$$

$$M = \text{Mean (मध्यमान)}$$

Sigma $\rightarrow \Sigma$ = जोड़ (the sum of)

X = प्राप्तांक (Score)

N = प्राप्तांकों की संख्या (No. of Scores)

मध्यांक (Median) - मध्यांक किसी श्रेणी के मध्य में स्थित वह बिन्दु है जो पूरी श्रेणी को दो समान भागों में बाँट देता है। अर्थात् इस बिन्दु अथवा पदमूल्य से 50 प्रतिशत पद मूल्य ऊपर होते हैं और 50 प्रतिशत पद मूल्य नीचे की ओर होते हैं। मध्यांक निकालने के लिए सम्पूर्ण श्रेणी को घटते (descending) या बढ़ते (ascending) क्रम में सजाना होता है।

मध्यांक निकालने की विधि-

अव्यवस्थित आँकड़ों का मध्यांक

$$Md = \left(\frac{N+1}{2}\right)^{\text{th}} \text{ item का मान}$$

N = कुल पदों की संख्या (No. of terms)

जैसे - 7, 8, 5, 9, 6, 4 ये अव्यवस्थित आँकड़ें हैं।

इसको बढ़ते क्रम में व्यवस्थित करना है:

4, 5, 6, 7, 8, 9

$$\text{अब } Md (\text{मध्यांक}) = \left(\frac{N+1}{2}\right) = \frac{6+1}{2} = \frac{7}{2} = 3.5 \text{ वें पद का मान}$$

उपरोक्त प्रश्न में तीसरी और चौथी संख्या का मध्यमान होगा। चूंकि तीसरी संख्या 6 और चौथी संख्या 7 है। और 6 और 7 का मध्यमान (Mean)

$$= \frac{6+7}{2} = \frac{13}{2} = 6.5$$

अतः अगुचित मध्यांक (Median) 6.5 है।

उपरोक्त प्रश्न में N सम (even) था किन्तु जब N विषम (odd) है तो मध्यांक (Median) इस प्रकार निकाला जायेगा:

Ex - 11, 17, 15, 18, 20, 19, 21

Ascending order - 11, 15, 17, 18, 19, 20, 21

$$\text{सूत्र } \rightarrow Md = \frac{N+1}{2} = \frac{7+1}{2} = \frac{8}{2} = 4$$

इसमें चौथी संख्या (18) मध्यांक (Median) होगी।

उपस्थित आँकों का मध्यक (Median)

$$\text{सूत्र - } Md(\text{Median}) = L + \frac{N/2 - \text{cumfb}}{f_m} \times s$$

यहाँ Md = मध्यक

L = उस वर्ग की शुरु, निम्नतम सीमा जिसमें मध्यक है
(Exact lower limit of that group in which median lies)

N/2 = कुल संख्या का आधा (One half of the No. of terms)

cumfb = जिस वर्ग में मध्यक है उस वर्ग से नीचे की संघटित आवृत्ति
(Cumulative frequency of the class in which median falls)

f_m = उस वर्ग की आवृत्ति जिसमें मध्यक है (frequency median)

s = वर्ग का आकार (Length of the class)

Example

C.I. (Class Interval)	(frequency) f	cumfb
90-94	2	56
85-89	2	54
80-84	4	52
75-79	8	48
70-74	6	40
<u>65-69</u>	11	34
60-64	09	23
55-59	07	14
50-54	05	07
45-49	00	02
40-44	02	02
N = 56		

$$34 - 11 = 23$$

मध्यक $\frac{N}{2} = 28$ वाँ पद, वर्ग 65-69 में पड़ेगा

अतः माध्यिका (Median) वर्ग 65-69 होगा

$$\text{सूत्र } Md = L + \frac{N/2 - \text{cumfb}}{f_m} \times s$$

$$64.5 + \frac{28 - 23}{11} \times 5$$

$$64.5 + \frac{5}{11} \times 5$$

$$= 64.5 + \frac{25}{11} = 64.5 + 2.27 = \underline{\underline{66.77}}$$

बहुलांक (Mode) → बहुलांक श्रेणी में उस पद या मूल्य का या माप होता है जिसकी आवृत्ति सबसे अधिक होती है। इसका अर्थ यह है कि बहुलांक पदमात्रा में ऐसा मूल्य होता है जो सबसे अधिक बार आता है। अर्थात् बहुलांक (Mode) वह मूल्य होता है जो बार-बार आता है और आवृत्ति (frequency) वितरण श्रेणी में यह वह बिन्दु या मूल्य होता है जिसकी अधिकतम आवृत्ति होती है। अंग्रेजी का Mode शब्द फ्रेंच भाषा के La Mode से बना है जिसका अर्थ होता है (Most fashionable).

6.

बहुलांक (Mode) का निर्धारण - यह दो प्रकार का होता है :

(i) अशुद्ध बहुलांक (Crude Mode) (ii) शुद्ध बहुलांक (True Mode)

अशुद्ध बहुलांक या अव्यवस्थित आँकड़ों का बहुलांक (Mode of Ungrouped data)

Ex- 5, 10, 6, 5, 7, 9, 8, 7, 6, 7

उपरोक्त आँकड़ों में 7 की आवृत्ति 3 बार हुई है अतः 7 अभीष्ट बहुलांक है (Mode)।

शुद्ध बहुलांक या व्यवस्थित आँकड़ों का बहुलांक (Mode of grouped data)

$$\text{Mode}(M_o) = 3 \text{ Median} - 2 \text{ Mean}$$

$$\text{बहुलांक} = 3 \text{ माध्यिका} - 2 \text{ माध्य}$$

$$\text{अर्थात् } M_o = L + \frac{f_a}{f_a + f_b} \times i$$

L = उस वर्ग की निम्नतम सीमा जिसमें आवृत्तियों की संख्या सबसे अधिक है (Limit of that class interval having highest frequency)

f_a = वर्ग के ऊपर वाली आवृत्ति (frequency above)

f_b = वर्ग के नीचे वाली आवृत्ति (frequency below)

i = वर्गान्तर का प्रकार (Length of class interval)

Example

C.I (Class Interval)	f (frequency)
50-54	3
45-49	2
40-44	3
35-39	4
30-34	8
25-29	6
20-24	2
15-19	1
10-14	1

$$M_o = L + \frac{f_a}{f_a + f_b} \times i$$

$$= 29.5 + \frac{4}{4+6} \times 5$$

$$= 29.5 + \frac{4}{10} \times 5 = 29.5 + \frac{20}{10} = 29.5 + 2 = \underline{\underline{31.5}}$$

—x—