महाराष्ट्र राज्य तंत्रशिक्षण मंडळ, मुंबई
(स्वायत्त) (ISO 9001:2015) (ISO/IEC 27001:2013)

अभियांत्रिकी आणि तंत्रज्ञान पदविका

शिक्षण पुस्तिका
(Learning Material)

STATISTICAL MODELLING FOR MACHINE LEARNING

313307

K-Scheme

स्टॅटिस्टिकल मॉडेलिन्ग फॉर मशीन लर्निंग

संगणक अभियांत्रिकी गट
(के–स्कीम)

मराठी-इंग्रजी (द्विभाषिक) माध्यम
(अभियांत्रिकी व तंत्रज्ञानातील तिसरे सत्र पदविका)

शिक्षण पुस्तिका

(Learning Material)

स्टॅटिस्टिकल मॉडेलिन्ग फॉर मशीन लर्निंग

Statistical Modelling for Machine Learning

(313307)

संगणक अभियांत्रिकी गट

(के-स्कीम)

मराठी-इंग्रजी (द्विभाषिक) मीडियम

(अभियांत्रिकी व तंत्रज्ञानातील तिसरे सत्र पदविका)

महाराष्ट्र राज्य तंत्रशिक्षण मंडळ, मुंबई

(स्वायत्त)(ISO 9001:2015) (ISO/IEC 27001:2013)

स्टॅटिस्टिकल मॉडेलिन्ग फॉर मशीन लर्निंग

Statistical Modelling for Machine Learning

(313307)

मार्गदर्शक

प्रा. विजय नामदेवराव कुकरे

विभागप्रमुख, संगणक अभियांत्रिकी

प्रकल्प समन्वयक

प्रा. विजय तुकाराम पाटील

विभागप्रमुख, संगणक अभियांत्रिकी

संकलक

प्रा. सतीश सावंत

अधिव्याख्याता, माहिती तंत्रज्ञान

प्रा. वृषाली पाटील

अधिव्याख्याता, संगणक अभियांत्रिकी

प्रा. अनुप्रिता बोरित्कार

अधिव्याख्याता, स्थापत्य अभियांत्रिकी

प्रा. स्मिता जाधव

अधिव्याख्याता, विज्ञान आणि मानविकी अभियांत्रिकी

प्रा. मेरलीन मॅथुस

अधिव्याख्याता, संगणक अभियांत्रिकी

महाराष्ट्र राज्य तंत्र शिक्षण मंडळ

(स्वायत्त) (ISO: ९००१:२०१५) (ISO/IES: २७००१-२०१३)

शासकीय तंत्रनिकेतन इमारत, चौथा मजला, ४९, खेरवाडी, बांद्रा (पूर्व), मुंबई - ४०० ०५१.

दूरध्वनी क्र.: ०२२-६२५४२१७० / १६१

Email : director@msbte.com Web : www.msbte.org.in

प्रास्ताविक

महाराष्ट्र राज्यातील पदविका स्तरावरील तंत्रशिक्षणामध्ये विद्यार्थ्यांचे रोजगार कौशल्य विकसित करून विद्यार्थ्यांचा सर्वांगीण विकास घडवून आणण्याकरिता महाराष्ट्र राज्य तंत्रशिक्षण मंडळ कटिबद्ध आहे. उद्योगधंद्यातील बदलत्या तंत्रज्ञानाशी संबंधित गरजा लक्षात घेऊन महाराष्ट्र राज्य तंत्र शिक्षण मंडळाकडून पदविका अभ्यासक्रम वेळोवेळी अद्यावत करण्यात येतो . अभियांत्रिकी पदविका अभ्यासक्रम शिकत असतांना संकल्पनात्मक ज्ञान, सुसंगत संदर्भ, प्रश्न विचारणे, विश्वसनिय पुरावे, कारणमीमांसा आणि सुस्पष्ट निकष यांचा वापर करून अर्थाची उकल करण्याची, विश्लेषण व मूल्यमापन करण्याची तसेच तर्काने अनुमान काढण्याची क्षमता म्हणजेच चिकित्सक विचार विद्यार्थ्यांमध्ये अधिक दृढ होतील असा मला विश्वास आहे. जेव्हा विद्यार्थी ज्ञान मिळवण्याच्या माध्यमाशी पूर्णपणे परिचित आणि सोयीस्कर असतात, तेव्हा त्यांच्यासाठी वर्गातील चर्चेत भाग घेणे सोपे होते, संकल्पनात्मक व सैद्धांतिक बाबींचे आकलन परिपूर्ण होते, संज्ञानात्मक क्षमता सुधारते आणि त्यांचा आत्मविश्वास देखील वाढतो या सर्व गोष्टींचा विचार करून मंडळाकडून शैक्षणिक सामुग्रीची निर्मिती करण्यात आलेली आहे . भारत देश हा खेड्यापाड्यातून विकसित झालेला देश असून ग्रामीण भागातील विद्यार्थ्यांना तांत्रिक शिक्षण घेतांना भाषेचा अडसर न येता तांत्रिक बाबींचा आशय समजून घेणे शक्य होईल या दृष्टिकोनातून महाराष्ट्र राज्य तंत्र शिक्षण मंडळाने पदविका स्तरावरील तांत्रिक शिक्षणाकरीता विद्यार्थ्यांना मराठी-इंग्रजी द्विभाषिक माध्यमाचा पर्याय शैक्षणिक वर्ष २०२१-२२ पासून उपलब्ध करून दिलेला आहे .

राष्ट्रीय शैक्षणिक धोरण-२०२० प्रादेशिक भाषेतील शिक्षणास प्रोत्साहन देते, ज्यामुळे विद्यार्थ्यांना तांत्रिक अभ्यासक्रमांसाठी प्रादेशिक भाषांतुन शिक्षणाचे माध्यम निवडता येते. सदर धोरणामुळे प्रादेशिक भाषांमध्ये तांत्रिक सामग्री आणि अभ्यास सामग्रीचा विकास आणि भाषांतर निर्माण करण्याची आवश्यकता आहे . त्यास अनुसरून मंडळाने मराठी-इंग्रजी द्विभाषिक माध्यमाचा पर्याय द्वितीय व तृतीय वर्षाकरिताही उपलब्ध करून देण्यात आला आहे. तसेच त्याकरिताची शैक्षणिक सामग्रीही संबंधीत भागधारकरांना उपलब्ध करून देण्यात येत आहे.

पदविका स्तरावरील तंत्रशिक्षण अधिक दर्जेदार करण्यासाठी महाराष्ट्रातील अनुभवी व तज्ञ अध्यापकांनी व्यवहारिक मराठी भाषा व इंग्रजी भाषेतील तांत्रिक शब्दावली यांचा वापर करून मराठी - इंग्रजी भाषेचा सुवर्णमध्य साधण्याचा प्रयत्न केलेला आहे. मंडळाच्या स्तरावर गठीत सुकाणू समितीमार्फत सदर शैक्षणिक सामुग्रीचा दर्जा, तसेच इतर बाबींची तपासणी करण्यात आलेली आहे. त्यामुळे सदर शैक्षणिक सामुग्री अधिक सम्पन्न झालेली असून विद्यार्थी त्यांच्या व्यक्तिमत्त्वाचा सुसंवादी आणि सर्वांगीण विकास साधतील. परिणामतः विश्वस्तरीय मनुष्यबळाच्या गरजा पूर्ण करण्यात महाराष्ट्र राज्य अग्रेसर राहील व पर्यायाने राष्ट्रनिर्मिती करीता निश्चितच हातभार लागेल , असा मला विश्वास आहे.

अभियांत्रिकी पदविका अभ्यासक्रमातील प्रमुख विषयांची मराठी-इंग्रजी द्विभाषिक शैक्षणिक सामुग्री बनविण्यासाठी अध्यापक व सुकाणू समितीचे सदस्य यांनी दर्शविलेले समर्पण व वचनबद्धता कौतुकास पात्र आहे , या सर्वांचे मी मनः पूवक अभिनंदन करतो !

(प्रमोद नाईक)
संचालक
म. रा. तंत्र शिक्षण मंडळ, मुंबई.

अनुक्रमणिका

अ. नु.	युनिटचे नाव	पृष्ठ क्रमांक
1	**स्टॅटिस्टिकल तंत्र** (Statistical Techniques)	1
2	**स्टॅटिस्टिकल मेथोड्स** (Statistical Methods)	47
3	**प्रोबॅबिलिटी ऑफ रँडम व्हेरिएबल्स** (Probability of Random Variable)	78
4	**इंटरपोलेशन** (Interpolation)	98
5	**सॅम्पलिंग मेथोड्स** (Sampling Methods)	114

युनिट-1
स्टॅटिस्टिकल तंत्र
(Statistical Techniques)

अभ्यासक्रम निष्पत्ती (Course Outcome):

CO1: दिलेल्या समस्यांचे निराकरण करण्यासाठी संभाव्यतेची तत्त्वे वापरा

घटक निष्पत्ती (Theory Learning Outcome):

1. फ्रेकेंव्सी वितरणावर आधारित समस्या सोडवा
2. सर्व प्रकारच्या डेटासाठी सरासरी, मध्य आणि मोडची गणना करा
3. ग्राफिकल पद्धत वापरून मोड आणि मीडियनशोधा
4. दिलेल्या डेटासाठी कार्ल पियर्सन आणि बॉली यांच्या स्क्युनेसचे सह-कार्यक्षमता शोधा
5. दिलेल्या डेटासाठी क्षणाच्या आधारावर कर्टोसिसचे मोजमाप मोजा

1.1 परिचय (Introduction)

सांख्यिकी ही विज्ञानाची एक शाखा आहे जी डेटाचे संकलन, वर्गीकरण किंवा आयोजन करते. संख्यात्मक डेटाचे विश्लेषण करून वैध निकष काढण्यासाठी आणि त्यानंतर अशा विश्लेषणाच्या आधारे वाजवी निर्णय घेण्याच्या पद्धतींशी संबंधित आहे.

इंडस्ट्री किंवा कॉमर्स, अर्थशास्त्र, शिक्षण, खगोलशास्त्र आणि संशोधन अभियांत्रिकी अशा सर्व क्षेत्रांमध्ये आकडेवारी खूप उपयुक्त आहे. सांख्यिकीय डेटा आणि सांख्यिकीय विश्लेषणाचे तंत्र मजुरी, किंमती, उपभोग, उत्पन्न आणि संपत्तीचे वितरण इत्यादीसारख्या विविध आर्थिक समस्यांचे निराकरण करण्यासाठी अत्यंत उपयुक्त ठरले आहे. डेटाचे संकलन आणि वर्गीकरण ही कोणत्याही सांख्यिकीय तपासणीची सुरुवात आहे.

व्हेरिएबल: (Variable): एक परिमाण जे एका व्यक्तीपासून दुसऱ्यामध्ये बदलते ते व्हेरिएबल म्हणून ओळखले जाते, व्हेरिएबल एकतर डिस्क्रिट किंवा सतत असू शकते.

उदाहरणार्थ: गुण मिळवणाऱ्या विद्यार्थ्यांचे 5-10, 10-15, 15-20, 20-25, इत्यादी गट केले असल्यास पहिल्या गटात 5-10 मध्ये ज्या विद्यार्थ्यांचे गुण 5 किंवा त्यापेक्षा जास्त आहेत परंतु 10 पेक्षा कमी आहेत अशा सर्व विद्यार्थ्यांचा समावेश करू. 10 गुण मिळविणाऱ्या विद्यार्थ्यांना पहिल्या गटात समाविष्ट केले जात नाही परंतु ते दुसऱ्या गटात म्हणजेच 10-15 मध्ये समाविष्ट केले जातात. अशा प्रकाराच्या वितरणास सतत वितरण म्हणतात. एका वर्गाची खालची मर्यादा आणि पुढच्या वर्गाची वरची मर्यादा यांच्यात अंतर नसल्याने त्याला सतत वितरण म्हणतात.

खंडित वितरणाचे निरंतर वितरणामध्ये रूपांतर करण्यासाठी, आम्ही द्वितीय श्रेणीची खालची मर्यादा आणि प्रथम वर्गची वरची मर्यादा यांचा फरक घेतो, त्यास 2 ने भागतो, सर्व खालच्या मर्यादांमधून परिणामी मूल्य वजा करतो आणि सर्व वरच्या मर्यादा मूल्य जोडतो.

उदाहरणार्थ: गट खालीलप्रमाणे असल्यास:

5-9, 10-14, 15-19, 20-24 येथे 10 आणि 9 चा फरक 1 आहे, 2 ने भागल्यास 0.5 मिळतात.

सतत वितरण 4.5-9.5, 9.5-14.5, 14.5-19.5, 19.5 आहे -24.5 इ.

क्युम्युलेटिव्ह फ्रेकेंव्सी (Cumulative Frequency): कधीकधी आपल्याला विशिष्ट मूल्यापेक्षा कमी संख्या आवश्यक असते, त्या मूल्यापर्यंत वर्गांची फ्रेकेंव्सी जोडतो आणि या संख्येला संचयी फ्रेकेंव्सी म्हणतो.

मीन, मोड, मीडियन (Mean, Mode, Median): फ्रेकेंव्सी वितरण काही मूल्याभोवती डेटाचे क्लस्टरिंग दर्शवते. अशा मूल्याला Central value केंद्रीय मूल्य आणि सरासरी म्हणतात. हे संपूर्ण डेटाचे सर्वाधिक प्रतिनिधी मूल्य देते.

केंद्रीय प्रवृत्तीचे सांख्यिकीय उपाय तीन प्रकारचे आहेत

(1) मीन (अंकगणित सरासरी) (Arithmetic Mean)

(२) मीडियन (Median)

(३) मोड (Mode)

(1) **अंकगणित मीन (Arithmetic mean):** हे डेटाचे सरासरी मूल्य आहे, म्हणजे सर्व वैयक्तिक निरीक्षणांची बेरीज करुन एकूण निरीक्षणांच्या संख्येने भागणे.

(2) **मोड (Mode):** मोड हे व्हेरिएबलचे मूल्य आहे ज्यामध्ये कमाल फ्रेकेंव्सी असते. दिलेल्या डेटाचे हे सर्वात सामान्य निरीक्षण आहे. त्याला मोडल व्हॅल्यू असेही म्हणतात.

(3) **मीडियन (Median):** जेव्हा डेटाची निरीक्षणे चढत्या किंवा उतरत्या क्रमाने मांडली जातात तेव्हा मीडियन हे मध्यवर्ती वस्तूचे मोजमाप असते. हे डेटा दोन समान भागांमध्ये विभाजित करते.

1.2 मीन(Mean), मीडियन(Median), मोड(Mode): गटबद्ध आणि गटबद्ध फ्रेकेंव्सी वितरण

(Classification of Data: Raw. Ungroup & Group data)

प्रकार 1.

* **(Raw Data)** कच्च्या डेटाचा मीन, मोड आणि मीडियन शोधण्यासाठी म्हणजे जेव्हा केवळ X1, X2,....., Xn निरीक्षणे दिली जातात.

प्रकार 2.

***Ungrouped Data** गटबद्ध न केलेल्या डेटाचा निरीक्षणे (X□) आणि फ्रेकेंव्सी (fi) मीन, मोड आणि मीडियन शोधण्यासाठी दिली जातात.

मूल्य	x_1	x_2	x_3	x_4
फ्रेकेंव्सी	f_1	f_2	f_3	f_4

प्रकार 3:

***Grouped Data** गटबद्ध डेटाचा मीन, मोड आणि मीडियन शोधण्यासाठी म्हणजे जेव्हा वर्ग-मांतर आणि फ्रेकेंव्सी दिली जाते.

वर्ग	C_1-C_2	C_2-C_3	C_3-C_4	C_{n-1}-C_n
फ्रेकेंव्सी	F_1	F_2	F_3	F_n

1.3 केंद्रीय प्रवृत्तीचे उपाय (Measures of Central Tendency):

प्रकार 1. (Raw Data)

1. Mean:

थेट पद्धत (Direct mean): जर X1, X2,..., Xn ही व्हेरिएबल x च्या निरीक्षणांची 'n' संख्या असेल तर

$$\text{Mean} = \frac{X_1+X_2+X_3+ \ldots\ldots\ldots + X_n}{Total\ No.of\ Observation} = \frac{\sum_{i=1}^{n} Xi}{n}$$

शॉर्ट कट मेथड (अस्युम्ड मीन मेथड): Shortcut Method (Assumed Mean Method) या पद्धतीमध्ये आपण मीनचे काही मूल्य गृहीत धरतो आणि नंतर मीनची वास्तविक गणना करतो.

$$\text{Mean} = A + \frac{\sum_{i=1}^{n} di}{n}$$

- A = Assumed Mean गृहीत धरलेले सरासरी (दिलेल्या डेटाचे अंदाजे केंद्र म्हणून घेतले पाहिजे)

- di = Xi-A (x चे विचलन; A पासून)
- n = एकूण संख्या निरीक्षणे.

2. मोड:

जर X1, X2,..., Xη दिलेली निरीक्षणांची 'n' संख्या असेल तर या कच्च्या डेटाचा मोड म्हणजे Xi चे मूल्य आहे; जे जास्तीत जास्त वेळा येते म्हणजेच सर्वात सामान्य मूल्य.

3. मीडियन:

जर X1, X2, Xn ही 'n' निरीक्षणे चढत्या क्रमाने मांडलेली असतील तर मीडियनहा क्रमाने मांडलेल्या डेटाचा मध्य असतो.

(i) जर n= विषम संख्याअसेल ,तर

मीडियन= $\frac{n+1}{2}$ व्या स्थानाचे निरीक्षण.

(ii) जर n= सम संख्या असेल तर

मीडियन= ($\frac{n}{2}$ व्या स्थानाचे निरीक्षण + $\frac{n+1}{2}$ व्या स्थानाचे निरीक्षण) / 2

प्रकार 2 (Ungrouped Data)

1. Mean

थेट पद्धत:(Direct method): जर x1, x2,......, X ही व्हेरिएबलची मूल्ये असतील x अनुक्रमे f1, f2, f फ्रिक्वेन्सीसह.

$$\text{मीडियन} = \frac{\sum fixi}{\sum fi} = \frac{\sum fixi}{N}$$

येथे, N =$\sum fi$ = एकूण निरीक्षणांची संख्या.

शॉर्ट कट मेथड (अस्युम्ड मीन मेथड): (Shortcut Method (Assumed Mean Method) (गृहीत धरलेली सरासरी पद्धत):

$$\text{मीन} = A+ \frac{\sum fidi}{\sum fi} = A+ \frac{\sum fixi}{N}$$

येथे,

- A = Assumed Mean गृहीत धरले आहे
- N = ∑fi = एकूण निरीक्षणे.
- Di = Xi-A, म्हणजे गृहित सरासरीमधून घेतलेल्या x च्या प्रत्येक मूल्याचे विचलन.

2. Mode: जर X1, X2,....., Xn ही अनुक्रमे f1, f2,… fn फ्रिक्वेन्सीसह व्हेरिएबल x ची मूल्ये असतील तर या वितरणाचा मोड म्हणजे कमाल फ्रेकेंव्सी असलेल्या 'x' चे मूल्य.

3. Median: येथे आपल्याला संचयी फ्रेकेंव्सी असलेली तक्ता तयार करायची आहे. मीडियनहे संचयी $\frac{N}{2}$ फ्रिकेंसीशी संबंधित X चे मूल्य आहे किंवा फ्रिक्वेन्सी पेक्षा फक्त मोठे आहे म्हणजे . $\frac{N+1}{2}$ जेथे N एकूण फ्रिक्वेन्सी आहे

प्रकार 3. (Grouped Data)

Class-interval	C_1-C_2	C_2-C_3	C_3-C_4	C_{n-1}-C_n
Frequency	F_1	F_2	F_3	F_n

Mean: या प्रकारच्या वितरणामध्ये, फरक एवढाच आहे की आपल्याला विविध वर्ग मध्यांतरांचे मध्यबिंदू विचारात घ्यावे लागतील.

थेट पद्धत-डायरेक्ट मेण (Direct mean):

$$\text{मीडियन} = \frac{\sum fixi}{\sum fi}$$

येथे Xi = (Mid Point) वर्ग मध्यांतराचा मध्य बिंदू म्हणजे

$$xi = \frac{[\text{वरची सीमा + खालची सीमा}]}{2}$$

येथे गणना करण्याचा अर्थ असा आहे की मध्य-बिंदू समान राहिल्याने अखंड वितरणाला सतत वितरणात रूपांतरित करण्याची गरज नाही.

शॉर्ट कट मेथड (अस्युम्ड मीन मेथड):Shortcut Method (Assumed Mean Method) (गृहीत धरलेली सरासरी पद्धत):

$$\text{मीडियन} = A + \frac{\sum fidi}{\sum fi} = A + \frac{\sum fixi}{N}$$

येथे,

- A = Assumed Mean गृहीत धरले आहे
- N = ∑fi = एकूण निरीक्षणे.
- Di = Xi-A, म्हणजे गृहित सरासरीमधून घेतलेल्या x च्या प्रत्येक मूल्याचे विचलन.

पायरी विचलन पद्धत- स्टेप डेव्हिएशन मेथोड (Step Deviation Method): आपण शॉर्ट कट पद्धत सोपी करू शकतो. गृहीत मध्यापासून घेतलेले सर्व विचलन C ने विभाजित केले आहेत, जेथे C वर्ग अंतराची रुंदी समान आहे.

अंकगणित मीन बाय स्टेप डेव्हिएशन पद्धतीने मोजण्यासाठी

$$\text{मीडियन} = A + \left(\frac{\sum fidi}{\sum fi} \times C\right) = A + \left(\frac{\sum fixi}{N} \times C\right)$$

येथे, C = वर्गांतर = वरची सीमा – खालची सीमा

येथे,

$$di = \frac{Xi - A}{C}$$

2.Mode :

$$\text{मोड} = L + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times C$$

येथे

- L= मोडल क्लासची खालची मर्यादा (Lower limit of modal class)
- f_0 = मोडल क्लासच्या वर्गाची वरची फ्रेकेंव्सी (Frequency of class previous to modal class).
- f_1= मोडल क्लासची फ्रेकेंव्सी किंवा कमाल फ्रेकेंव्सी. (Frequency of modal class or Maximum frequency).
- f_2 = मोडल क्लासच्या वर्गाची खालची फ्रेकेंव्सी (Frequency of modal class).
- C= वर्ग अंतराची रुंदी. (Width of class interval)
- टीप: मोडल क्लास हा जास्तीत जास्त फ्रेकेंव्सी असलेला वर्ग आहे. फ्रेकेंव्सी वितरण मध्ये दिले आहे

3 Median: जेव्हा वर्ग अंतराच्या अटी,

$$\text{मीडियन} = L + \left[\frac{\left(\frac{N}{2}\right) - f_c}{f_m} \; X \; C\right]$$

- L = मीडियनवर्गाची खालची मर्यादा. (lower limit of median class)
- N = एकूण फ्रेकेंव्सी (Total frequency)
- Fm= फ्रेकेंव्सी मोडल क्लासची फ्रेकेंव्सी किंवा कमाल फ्रेकेंव्सी. (Frequency of median class)
- C= वर्ग अंतराची रुंदी. (Width of class interval)
- f_c = मागील वर्ग ते मध्य वर्गाची संचयी फ्रेकेंव्सी. (Cumulative Frequency of previous class to median class)

टीप: मध्यवर्ती वर्ग हा संचयी फ्रेकेंव्सी N/2 किंवा N / 2 पेक्षा मोठा वर्ग आहे

उदाहरणांसह स्पष्टीकरण

मध्य(Mean): अंकगणित सरासरी मूल्य संचाच्या सदस्यांची मूल्ये एकत्र जोडून आणि संचातील सदस्यांची संख्या यांचा भागाकार करून शोधले जाते. अशा प्रकारे,

सर्वसाधारणपणे, संचाचा मध्य: {x1, x2, x3,..., xn} आहे

$$x = \frac{x1 + x2 + x3 + \cdots + xn}{n}$$

ग्रीक अक्षर 'सिग्मा' $\sum$ आहे आणि याचा अर्थ 'बेरीज' आहे आणि $\bar{x}$ (x-bar म्हणतात) हे सरासरी मूल्य दर्शविण्यासाठी वापरले जाते.

संचाचा मध्य संख्यांची: {4,5,6,9} आहे:

$$\bar{x} = \frac{4+5+6+9}{4} \text{ ,i.e.6}$$

मीडियन (Median):

(a) परिमाणाच्या चढत्या क्रमाने सेट रँकिंग करणे, आणि

(b) विषम संख्या असलेले सेटसाठी मधल्या सदस्याचे मूल्य निवडणे सदस्यांची, किंवा सदस्यांची सम संख्या असलेल्या संचांसाठी दोन मीडियन सदस्यांच्या सरासरीचे मूल्य शोधणे.

उदाहरण ,

संच: {7,5,74,10} असे रँक केले आहे {5,7,10,74}, आणि त्यात सम संख्या असल्यामुळे (या प्रकरणात चार), सदस्य 7 आणि 10 चा मध्य घेतले आहे, 8.5 चे मीडियनमूल्य देते.

त्याचप्रमाणे, संच: {3,81,15,7,14} हे {3,7,14,15,81} म्हणून रँक केले आहे आणि मध्यवर्ती मूल्य 14 हे मीडियन सदस्याचे मूल्य आहे.

मोड (Mode): मॉडेल मूल्य, किंवा मोड, सर्वात सामान्यपणे आहे सेटमध्ये येणारे मूल्य. जर दोन व्हॅल्यू यासह आढळतात समान फ्रेकेंव्सी, संच 'द्वि-मॉडल' आहे.

संच:{5,6,8,2,5,4,6,5,3} चे मॉडेल मूल्य 5 आहे, कारण 5 चे मूल्य असलेला सदस्य तीन वेळा येतो.

उदाहरण -1.1: सेटसाठी मध्य, मीडियनआणि मोड निश्चित करा:

{2,3,7,5,5,13,1,7,4,8,3,4,3}

$$\bar{x} = \frac{2 + 3 + 7 + 5 + 5 + 13 + 1 + 7 + 4 + 8 + 3 + 4 + 3}{13} = \frac{65}{13} = 5$$

सेट रँकिंग :

{1,2,3,3,3,4,4,5,5,7,7,8,13} करून मधला मूल्य म्हणजे सातवा सदस्य, म्हणजे 4, अशा प्रकारे मध्यवर्ती मूल्य 4 आहे.

मोडल मूल्य हे चे मूल्य आहे सर्वात सामान्यपणे आढळणारा सदस्य आणि 3 आहे, जे तीन वेळा उद्भवते, इतर सर्व सदस्य फक्त एक किंवा दोनदा घडतात.

उदाहरण-1.2: बातमी विक्रेत्याने सहा दिवसांसाठी घेतलेल्या पाउंडमधील पैसे खालील डेटा संच संदर्भित करते. मध्य, मीडियनआणि मॉडेल मूल्ये निश्चित करा:
{27.90,34.70,54.40,18.92,47.60,39.68}

(Mean Value) सरासरी मूल्य $= \frac{27.90 + 34.70 + 54.40 + 18.92 + 47.60 + 39.68}{6} = 37.20$

रँक केलेला संच आहे: {18.92,27.90,34.70,39.68,47.60,54.40}
संच सदस्य संख्या सम संख्या असल्याने, सरासरी मधल्या दोन सदस्यांपैकी मीडियनदेण्यासाठी घेतले जाते मूल्य, म्हणजे

(Median value) मीडियनमूल्य $= \frac{34.70 + 39.68}{2} = 37.19$

कोणत्याही दोन सदस्यांना समान मूल्य नसल्यामुळे, या संचाला मोड नाहीआहे.

उदाहरण 1 ते 4 मध्ये, दिलेल्या सेटसाठी मध्य, मीडियनआणि मॉडेल मूल्ये ठरवा

1. {3, 8, 10, 7, 5, 14, 2, 9, 8}
2. {26, 31, 21, 29, 32, 26, 25, 28}
3. {4.72, 4.71,4.74 ,4.73, 4.72 4.71,4.73 ,4.72}
4. {73.8, 126.4, 40.7, 141.7, 28.5, 237.4, 157.9}

उदाहरण-1.3: फ्रेकेंव्सी वितरण 48 प्रतिरोधकांच्या ओममधील प्रतिकाराचे मूल्य दाखवले आहे. प्रतिकाराचे सरासरी मूल्य निश्चित करा.

C.I.	20.5–20.9	21.0–21.4	21.5. 21.9	22.0–22.4	22.5-22.9	23.0-23.4
Fi	3	10	11	13	9	2

C.I.	20.5–20.9	21.0–21.4	21.5.–21.9	22.0–22.4	22.5-22.9	23.0-23.4
Xi	20.7	21.2	21.7	22.2	22.7	23.2
Fi	3	10	11	13	9	2

गटबद्ध डेटासाठी, सरासरी मूल्य द्वारे दिले जाते:

$x = \frac{\sum(f(x))}{\sum f}$

जेथे f वर्ग फ्रेकेंव्सी आहे आणि x हे वर्ग मध्य बिंदू मूल्य आहे. म्हणून सरासरी मूल्य,

$x = \frac{(3 \times 20.7) + (10 \times 21.2) + (11 \times 21.7) + (13 \times 22.2) + (9 \times 22.7) + (2 \times 23.2)}{48} = \frac{1052.1}{48} = 21.919$

म्हणजे सरासरी मूल्य 21.9 आहे.

उदाहरण-1.4: खालील तक्त्यामध्ये 10वीच्या 100 विद्यार्थ्यांच्या परीक्षेत मिळालेल्या गुणांचे फ्रेकेंव्सी वितरण दाखवले आहे. मीडियन गुण शोधा.

परीक्षेतील गुण	0-20	20-40	40-60	60-80	80-100
विद्यार्थ्यांची संख्या	4	20	30	30	6

उत्तर: N = 100, N/2 = 50

C.I.	f	c.f

0-20	4	4
20-40	20	24
40-60	30	54
60-80	40	94
80-100	6	100

मीडियन वर्ग

= N/2 th मूल्य

= 50th मूल्य

= 50

मीडियन वर्ग 40-60 आहे

Here L=40, h=20, f=30, c.f. =24

Median= L + $\frac{h}{f}$ x (N/2 –c.f.)

= 40 + $\frac{20}{30}$ x (50 – 24.)

= 40 + 17.33

= 57.33

उदाहरण-1.5: खालील फ्रिक्केन्सी वितरण सारणीवरून मीडियनशोधा.

पेक्षा कमी वय (वर्षे)	10	20	30	40	50	60
व्यक्तींची संख्या	3	10	22	40	54	71

उत्तर:

Age less than(yrs)	**Class**	**No of persons(f)**	**c.f.**
10	0-10	3	3
20	10-20	10	7
30	20-30	22	12
40	30-40	40	18
50	40-50	54	14
60	50-60	71	17

Here N=71,N/2=71/2 =35.5 ,h=10

मीडियन वर्ग 30-40 आहे

Here L=30, h=10, f=30, c.f. =22

Median= L + $\frac{h}{f}$ x (N/2 –c.f.)

= 30 + $\frac{10}{18}$ x (35.5 – 22.)

= 30+7.5

=37.5

व्यक्तीचे मीडियनवय =37.5 वर्षे

1.4 Concept of Quartiles, decile & percentiles

- विभाजन मूल्ये-चतुर्थांश, डेसील, टक्केवारी.(Partial Values- Quartiles ,Deciles, Percentiles)
- विभाजन मूल्यांचे ग्राफिकल स्थान

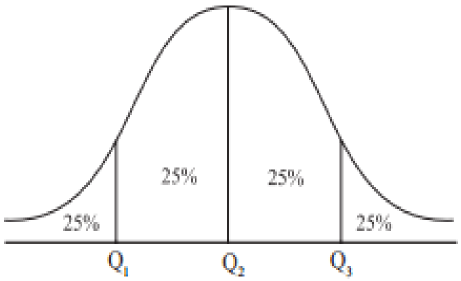
परिचय (Introduction): एका विशिष्ट शाळेला त्यांच्या मागील वर्गातील गुणांच्या अधीन 200 विद्यार्थ्यांचे समान संख्याबळाचे चार विभाग करण्यात स्वारस्य आहे. परिस्थितीनुसार आवश्यक संख्येच्या समान भागांमध्ये डेटा विभाजित करणे. डेटाला समान भागांमध्ये विभाजित करण्याच्या या प्रक्रियेस विभाजन म्हणतात आणि डेटाला आवश्यक संख्येच्या समान भागांमध्ये विभाजित करणार्या मूल्यांना विभाजन मूल्य म्हणतात.

1.4.1 विभाजन मूल्ये.(Partial Values): समजा आपल्याकडे 15 विद्यार्थ्यांच्या गणितातील गुणांचा खालील डेटा आहे. 23,21,26,27,20,18,22,13,14,25,34,28,35,40,38.
निरिक्षणांची त्यांच्या विशालतेच्या चढत्या क्रमाने मांडणी करा.

13,14,18,20,21,22,23,25,26,27,28,34,35,38,40.

निरीक्षण 20, 25, 34 च्या स्थानांबद्दल काय म्हणता येईल?
20 च्या खाली 3 निरीक्षणे, 20 आणि 25 मधील 3 निरीक्षणे, 25 आणि 34 मधील 3 निरीक्षणे आणि 34 वरील 3 निरीक्षणे आहेत. ही मूल्ये डेटाला 4 समान भागांमध्ये विभाजित करतात, म्हणून त्यांना विभाजन मूल्ये म्हणतात.

Definition (Partial values):जेव्हा निरीक्षणे त्यांच्या परिमाणाच्या चढत्या क्रमाने मांडली जातात, आणि पुढे, त्यांना 'n' समान भागांमध्ये (n=2,3,...) विभागले जातात, तेव्हा विभाजनाच्या स्थितीत असलेल्या निरीक्षणांना विभाजन मूल्ये म्हणतात.
विभाजन मूल्यांपैकी एक म्हणजे मीडियनया संकल्पनेचा आम्ही आधीच अभ्यास केला आहे. मीडियनडेटा दोन समान भागांमध्ये विभागतो. मीडियनहे विभाजन मूल्याचा एक विशेष प्रकार आहे कारण ते विभाजित करते. आपण चतुर्थक, डेसील आणि पर्सेंटाइल (Quartiles, Deciles, Percentiles) तीन प्रकारच्या विभाजन मूल्यांचा अभ्यास करणार आहोत.
चतुर्थांश (Quartiles): चतुर्थांश निरीक्षणांना 4 मध्ये विभाजित करतात चढत्या क्रमाने व्यवस्था केल्यावर समान भाग . Q1, Q2 आणि Q3 असे तीन चतुर्थांश (Quartiles) आहेत.



- Q_1 हा पहिला किंवा खालचा चतुर्थांश आहे. 25% निरीक्षणे Q_1 च्या खाली आहेत आणि 75% Q1 वर आहेत.
- Q_2 हा दुसरा चतुर्थांश (मीडियन) आहे जो डेटाला दोन समान भागांमध्ये विभाजित करतो 50% निरीक्षणे वर आणि 50% खाली आहेत
- Q_3 हा तिसरा किंवा वरचा चतुर्थांश आहे. 75% निरीक्षणे Q_3 च्या खाली आहेत आणि 25% Q_3 वर आहेत.

कच्च्या डेटासाठी क्वार्टाइलचे सूत्र खालीलप्रमाणे आहे:

$$i^{th} \text{ Quartile} = Q_1 = \text{value of } i\,\frac{n+1}{4} \text{ th observation. } i=1,2,3$$

when n= total no of observation.
चतुर्थकाचे(Quartile) मूल्य खालीलप्रमाणे मोजले जाते:

ith चतुर्थांश = 6व्या निरीक्षणाचे मूल्य + 0.25 (7व्या निरीक्षणाचे मूल्य – 6व्या निरीक्षणाचे मूल्य)
गटबद्ध (grouped data) डेटासाठी चतुर्थकांचे सूत्र खालीलप्रमाणे आहे:

$$Q_1 = L + \frac{h}{f} \times \left(\frac{iN}{4} - c.f.\right),\ i = 1, 2, 3$$

जेथे

- L = ith चतुर्थांशाची खालची सीमा वर्ग (lower boundary of *i*th quartile class)
- h = ith चतुर्थक वर्गाची वर्ग रुंदी(class width of *i*th quartile class)
- f = ith चतुर्थक वर्गाची फ्रेकेंव्सी (frequency of *i*th quartile class)
- c.f.= चतुर्थक वर्गाच्या अगदी आधीच्या वर्गाच्या संचयी वारंवारतेपेक्षा कमी less than cumulative frequency of the class just preceding *i*th quartile class)
- N = एकूण फ्रेकेंव्सी. (total frequency.)

उदाहरण 1.6 : 19 विद्यार्थ्यांचे गुण खाली दिले आहेत:

41, 21, 38, 27, 31, 45, 23, 26, 29, 30, 28, 25, 35, 42, 47, 50, 29, 31, 35

वरील डेटासाठी सर्व चतुर्थांशांची गणना करा.

उत्तर: प्रथम खालीलप्रमाणे चढत्या क्रमाने डेटाची मांडणी करा:

21, 23, 25, 26, 27, 28, 29, 29, 30, 31, 31, 35,35, 38, 41, 42, 45, 47, 50.

येथे , n=19

Q_1= value of $\frac{n+1}{4}\ th\ observation$

= $\frac{19+1}{4}\ th\ observation$

= value of 5th observation

Q_1=27

Q_2 = value of $2(\frac{n+1}{4})th\ observation$

=$2(\frac{19+1}{4})\ th\ observation$

= value of 10th observation

Q_2 =31

Q_3 = value of$3(\frac{n+1}{4})\ th\ observation$

=$3(\frac{19+1}{4})\ th\ observation$

= value of 15th observation

Q_3=41

उदाहरण-1.7: 12 कामगारांच्या दैनंदिन वेतनासाठी (rs) चतुर्थांशांची गणना करा:

200,280,310,180,190,170, 320,330,220,210,380,400.

उत्तर: प्रथम खालीलप्रमाणे चढत्या क्रमाने डेटाची मांडणी करा:

170,180,190,200,210,220,280,310,320,330,380,400

येथे, n = 12

Q_1 = value of $\frac{n+1}{4}\ th\ observation$

= $\frac{12+1}{4}\ th\ observation$

= value of 3.25th observation

= value of 3^{th} observation + 0.25 (value of 4^{th} observation- value of 3^{th} observation)

= 190+0.25(200-190)

Q_1 = 192.5

Q_2 = value of $2(\frac{n+1}{4})th\ observation$

$=2(\frac{12+1}{4})\ th\ observation$

= value of 6.5th observation

= value of 6th observation + 0.5 (value of 7th observation value of 6th observation -)

= 220+0. 5(280-220)

= 220+0.5 (60)

Q2 = 250

Q3 = value of$3(\frac{n+1}{4})\ th\ observation$

$=3(\frac{12+1}{4})\ th\ observation$

= value of 9.75th observation

= value of 9th observation + 0.75 (value of 10th observation- value of 9th observation)

= 320+0. 75(330-320)

= 320+0.75 (10)

Q3 = 327.5

उदाहरण-1.8: खालील डेटासाठी चतुर्थांशांची गणना करा:

उंची (इंच मध्ये)	4.8	4.9	5.0	5.1	5.2	5.3	5.4	5.5	5.6	5.7
विद्यार्थ्यांची संख्या	5	7	13	16	25	14	9	6	3	2

उत्तर:

संचयी फ्रेकेंव्सी सारणी खालीलप्रमाणे तयार करा:

Height (in inches)	No of students(f)	Less than cumulative frequency
4.8	5	5
4.9	7	12
5.0	13	25
5.1	16	41 (Q_1)
5.2	25	66 (Q_2)
5.3	14	80 (Q_3)
5.4	9	89
5.5	6	95
5.6	3	98
5.7	2	100
Total	100	

येथे, n = 100

By comparing i $\frac{n+1}{4}$ th with cumulative frequency (c.f.)

Q_1 = value of $\frac{n+1}{4}\ th\ observation$

$= \frac{100+1}{4}\ th\ observation$

= value of 25.25^{th} observation

= 25^{th} observation + 0.25 (26^{th} observation-25^{th} observation)

= 5+0.25(5.1-5)

Q_1 = 5.025

Q_2 = value of $2(\frac{n+1}{4}) th\ observation$

$=2(\frac{100+1}{4})\ th\ observation$

= value of 50.5^{th} observation

Q_2 = 5.2

Q_3 = value of$3(\frac{n+1}{4})\ th\ observation$

$=3(\frac{100+1}{4})\ th\ observation$

= value of 75.75^{th} observation

Q_3 = 5.3

उदाहरण-1.9: महामार्ग पोलीस विभागाने सर्वेक्षण केले आणि संख्येचा वेग मोजला. खालील महामार्गावरील गाड्या वितरण प्राप्त झाले

Speed below (in kms/hours)	65	70	75	80	85	90
No of cars	19	44	99	184	194	200

75% च्या खाली कारचा वेग असलेल्या गतीची गणना किमी/तास मध्ये करा, म्हणजेच आपल्याला Q3 चे मूल्य काढावे लागेल.

उत्तर: प्रथम, वर्ग तयार करा आणि फ्रेकेंव्सी सारणी खालीलप्रमाणे:

Speed below (in kms/hours)	No of cars	Less than cumulative frequency
Below 65	19	19
65-70	44-19=34	44
70-75	99-44=55	99
75-80	184-99=85	184 (Q_3)
80-85	194-184=10	194
85-90	200-194=6	200

येथे, N = 200

By comparing $3\ \frac{N}{4}$ = 150

Q_3 वर्गात 75-80 आहे

Here L=75, h=5, f=85, c.f. =99

$Q_3 = L + \frac{h}{f}$ x (3N/4 –c.f.)

$= 75 + \frac{5}{85}$ x (99)

= 75+3

=78

75% गाड्या महामार्गावरून ताशी 78 किमी पेक्षा कमी वेगाने जात आहेत.

टीप:

1. विभाजन मूल्यांसाठी खंडित वर्गांना सतत स्वरूपात रूपांतरित करणे अनिवार्य नाही. उत्तर काही दशांश आकृतीनुसार भिन्न असू शकते, जे अद्याप बरोबर आहे.
2. गहाळ फ्रेकेंव्सी मूल्य दशांश मध्ये असल्यास, ते जवळच्या पूर्ण संख्येच्या अंदाजे करा.

उदाहरण-1.10: Q2 चे खालील फ्रेकेंव्सी वितरण मूल्य 22 आहे, गहाळ फ्रेकेंव्सी शोधा.

class	0.5-9.5	10.5-19.5	20.5-29.5	30.5-39.5	40.5-49.5
frequency	5	8	?	4	3

उत्तर: आपण वर्ग सतत(continuous) करू आणि गृहीत धरू की x ही मिसिंग फ्रिक्वेन्सी आहे. संचयी फ्रेकेंव्सी सारणी खालीलप्रमाणे दिली आहे:

class	frequency	Less than cumulative frequency
0-10	5	5
10-20	8	13
20-30	X	13+x
30-40	4	17+x
40-50	3	20+x

येथे, 2nd quartile (Q_2) = 22

By comparing $2\frac{N}{4}$ = 22

Q_2 वर्गात 20-30 आहे

Here L=20, h=10, f=x, c.f. =13

$Q_2 = L + \frac{h}{f} \times (N/2 - c.f.)$

$22 = 20 + \frac{10}{x} \times \left(\frac{20+x}{2} - 13\right)$

$-20 = \frac{10}{x} \times \left(\frac{20+x}{2} - 13\right)$

$2x = 5 \times (20 + x - 26)$

$2x = 100 + 5x - 130$

$3x = 30$

$X = 10$

म्हणून, गहाळ Missing frequency फ्रेकेंव्सी 10 आहे.

सराव उदाहरण (Exercise)

1. खालील निरीक्षणांच्या मालिकेसाठी सर्व चतुर्थकांची गणना करा:

16, 14.9, 11.5, 11.8, 11.1, 14.5, 14, 12, 10.9, 10.7, 10.6, 10.5, 13.5, 13, 12.6.

2. दहा 10 विद्यार्थ्यांची उंची (सेमी. मध्ये) खाली दिली आहे:

148, 171, 158, 151, 154, 159, 152, 163, 171,145.

वरील डेटासाठी Q1 आणि Q3 ची गणना करा.

3. एका विशिष्ट परिसरातील कुटुंबांची विजेचा मासिक वापर खाली दिले आहे:

205, 201, 190, 188, 194, 172, 210, 225, 215,232, 260, 230.

25% च्या खाली असलेल्या कुटुंबांची विजेच्या वापराची गणना (युनिटमध्ये) करा.

4. खाली कुटुंबांचा दैनंदिन खर्च समावेश आहे. 75% च्या खालील डेटासाठी कुटुंबांची खर्चाची गणना (रु. मध्ये)करा.

दैनंदिन खर्च (रु. मध्ये)	350	450	550	650	750
कुटुंबांची संख्या	16	19	24	28	13

5. खालील फ्रेकेंव्सी वितरण मध्ये सर्व चतुर्थकांची गणना करा:

ई-व्यवहाराची संख्या (दररोज)	0	1	2	3	4	5	6	7
दिवसांची संख्या	10	35	45	95	64	32	10	9

6. कारखान्यातील 200 पुरुष प्रौढांच्या उंचीचे फ्रेकेंव्सी वितरण खालीलप्रमाणे आहे:

सेमी मध्ये उंची	पुरुष प्रौढांची संख्या
145-150	4
150-155	6
155-160	25
160-165	57
165-170	64
170-175	30
175-180	8
180-185	6

मध्यवर्ती उंची शोधा.

7. एका वर्गातील 50 विद्यार्थ्यांच्या दर आठवड्याला होणाऱ्या खिशातील खर्चाचा डेटा खालीलप्रमाणे आहे. हे ज्ञात आहे की वितरणाचा मीडियन120 आहे. गहाळ फ्रिक्वेन्सी शोधा.

Wages more than (in Rs)	No of workers
8000	160
9000	155
10000	137
11000	103
12000	57
13000	23
14000	10
15000	1
16000	0

8. एका विशिष्ट बँकेतील 100 ज्येष्ठ नागरिकांच्या मुदत ठेवींच्या कालावधीसाठी खालील गटबद्ध डेटा आहे:

Fixed deposits (In days)	0-180	180-360	360-540	540-720	720-900
No of senior citizens)	15	20	25	30	10

केंद्रीय 50% ज्येष्ठ नागरिकांच्या मुदत ठेवींच्या मर्यादांची गणना करा.

9. वितरणाचा मीडियन1504 आहे असे दिलेली नसलेली फ्रेकेंव्सी शोधा.

Life in hours	950-1150	1150-1350	1350-1550	1550-1750	1750-1950	1950-2150
No of bulbs	20	43	100	-	23	13

Deciles: Deciles D1, D2, ... D9 द्वारे दर्शविले जातात.
कच्च्या डेटासाठी डेसिल्सचे सूत्र खाली दिले आहे:
i^{th} Decile = D_i
=value of observation i $(\frac{n+1}{10})$ th
i = 1, 2,..., 9
गटबद्ध डेटासाठी डेसिल्सचे सूत्र आहे,

$$D = L + \frac{h}{f} (\frac{iN}{100} - c.f.),\ \ i = 1, 2,...., 99$$

येथे

- L = ith decile वर्गाची खालची सीमा
- h = ith decile वर्गाची वर्ग रुंदी
- f = डेसिल क्लासची फ्रेकेंव्सी
- c.f.= ith decile वर्गाच्या आधीच्या वर्गाच्या संचयी वारंवारतेपेक्षा कमी
- N = एकूण फ्रेकेंव्सी

टक्केवारी (Percentiles): जेव्हा परिमाणाच्या त्यांची चढत्या क्रमाने मांडणी केली जाते, तेंव्हा टक्केवारी ही मूल्य निरीक्षणांना 100 समान भागांमध्ये विभाजित करतात. टक्केवारी P_1, P_2, ...,P_{99} द्वारे दर्शविली जाते. कच्च्या डेटासाठी पर्सेंटाइलचे सूत्र खाली दिले आहे:
i^{th} Perecentile = P_i
=value of observation i $(\frac{n+1}{100})$ th
i = 1, 2,... ,99
गटबद्ध डेटासाठी पर्सेन्टाईलचे सूत्र आहे,

$$D = L + \frac{h}{f} (\frac{iN}{100} - cf)\ ,\ \ i = 1, 2,...., 99$$

येथे,

- L = ith टक्केवारीची खालची सीमा
- h = ith टक्केवारी वर्गाची वर्ग रुंदी
- f = ith टक्केवारी वर्गाची फ्रेकेंव्सी
- c.f = संचयी वारंवारतेपेक्षा कमी
- N = एकूण फ्रेकेंव्सी

व्याख्येवरून असे दिसून येते की,
(1) Q2 = D5 = P50
(2) Q1 = P25
(3) Q3 = P75

उदाहरण-1.11: खालील डेटा साठी 3रा डेसील आणि 70वा पर्सेंटाइल मोजा:
841,289,325,225,784,729,625,400,324,169.

उत्तर: प्रथम, आम्ही डेटाची व्यवस्था करतो खालीलप्रमाणे चढत्या क्रमाने:
169, 225, 289, 324, 325, 400, 625, 729, 784,841
येथे, n = 10

D_3 = value of $3\frac{n+1}{10}$ $th\ observation$

= $3\,\frac{10+1}{10}$ $th\ observation$

= value of 3.3rd observation

= 3rd observation + 0.3 (4th observation-3rd observation)

= 289+0.3(324-289)

D_3 = 299.5

P_{70} = value of $70\frac{n+1}{100}$ $th\ observation$

= $70\,\frac{10+1}{100}$ $th\ observation$

= value of 7.7th observation

= 7th observation + 0.7 (8th observation-7th observation)

= 625+0.7(729-625)

P_{70} = 697.8

उदाहरण-1.12: खालील डेटासाठी D4 आणि P55 शोधा:

सदोष उत्पादनांची संख्या	30	35	40	45	50	55	60
कंपनीची संख्या	12	35	10	15	8	7	8

उत्तर: खालीलप्रमाणे संचयी फ्रेकेंव्सी सारणीपेक्षा कमी तयार करा:

सदोष उत्पादनांची संख्या	कंपनीची संख्या	संचयी वारंवारतेपेक्षा कमी
30	12	12
35	35	47 (D_4)
40	10	57 (P_{55})
45	15	72
50	8	80
55	7	87
60	8	95

प्रथम, आम्ही डेटामध्ये व्यवस्था करतो खालीलप्रमाणे चढत्या क्रमाने:

n = 95

D_4 = value of $4\frac{n+1}{10}$ $th\ observation$

= $4\,\frac{95+1}{10}$ $th\ observation$

= value of 38.4th observation

D_4 = 35

P_{55} = value of $55\frac{n+1}{100}$ $th\ observation$

= $55\,\frac{95+1}{100}$ $th\ observation$

= value of 52.8th observation

P_{55} = 40

उदाहरण-1.13: खालील डेटावरून 4 था डेसील आणि 21 व्या पर्सेंटाइलची गणना करा:

नफा (लाख रु. मध्ये)	0.5-4.5	5.5-9.5	10.5-14.5	15.5-19.5	20.5-24.5
कंपनीची संख्या	7	18	25	30	20

उत्तर: वर्ग सतत करा आणि खालीलप्रमाणे संचयी फ्रेकेंव्सी सारणी तयार करा:

नफा (लाख रु. मध्ये)	कंपनीची संख्या	संचयी वारंवारतेपेक्षा कमी
0-5	7	7
5-10	18	25 (P_{21})
10-15	25	50 (D_4)
15-20	30	80
20-25	20	100

D_4 = value of $4\frac{N}{10}$ $th\ observation$

= $4\frac{100}{10}$ $th\ observation$

= value of 40^{th} observation

D_4 वर्गात 10-15 आहे

Here L=10, h=100, f=18, c.f. =25

Median = L + $\frac{h}{f}$ x ($\frac{4N}{10}$ –c.f.)

= 10 + $\frac{5}{25}$ x ($\frac{4\ X\ 100}{10}$ – 25.)

= 10 + $\frac{5}{25}$ x (40 – 25)

= 10 + 3

D_4 = 13

Here L=5, h=5, f=18, c.f. =25

Median= L + $\frac{h}{f}$ x ($\frac{21N}{100}$ –c.f.)

= 5 + $\frac{5}{18}$ x ($\frac{21\ X\ 100}{100}$ – 25.)

= 5 + $\frac{5}{18}$ x (21 – 7)

= 5 + $\frac{70}{18}$

= 5+3.89

P_{21} = 8.89

For percentile

P_{21} = value of $\frac{21(100)}{100}$ observation

value of $21th$ observation

P_{21} वर्गात 5-10 आहे

Here L=5, h=5, f=18, c.f. =7

Median = L + $\frac{h}{f}$ x ($\frac{21N}{100}$ –c.f.)

= 5 + $\frac{5}{18}$ x ($\frac{21\ X\ 100}{100}$ – 7)

= 5+ $\frac{5}{18}$ x (21 – 7)

$= 5 + 3.89$

$P_{21} = 8.89$

उदाहरण-1.14: विविध विभागांमार्फत कार्यालयीन फायलींच्या प्रवासाशी संबंधित अभ्यास करण्यात आला. आरंभ विभागाकडे परत येण्यासाठी फाईलने घेतलेल्या वेळेचे (महिन्यांमध्ये) वितरण खाली दिले आहे:

Time	1-2	2-3	3-4	4-5	5-6	6 & above
No of files	30	45	40	50	6	4

शोधा:

(a) 75% फायली ज्या वेळेनुसार

(b) ज्या वेळेत जास्तीत जास्त ६०% फाइल्स परत येतात.

उत्तर: संचयी फ्रेकेंव्सी सारणी खालीलप्रमाणे तयार करा:

Times (In months)	No of files (f)	Less than c.f
1-2	30	30
2-3	45	75
3-4	40	115 (P_{60})
4-5	50	165 (P_{75})
5-6	6	171
6 & above	4	175

75% फाईल्स किती वेळात परत येतात हे शोधण्यासाठी म्हणजे P75 चे मूल्य शोधावे लागेल आणि जास्तीत जास्त 60% फाईल्स परत येण्याची वेळ आम्हाला शोधावी लागेल.

म्हणजे आपल्याला P_{75} शोधायचा आहे

P_{75} = value of $=\frac{75(175)}{100}$ observation

value of $=131.25th$ observation

P_{75} वर्गात 4-5 आहे

Here L=4, h=1, f=50, c.f. =115 ,N=175

$P_{75} = L + \frac{h}{f} \text{ x } (\frac{75N}{100} - \text{c.f.})$

$= 4 + \frac{1}{50} \text{ x } (\frac{75 X 175}{100} - 115)$

$= 4 + \frac{1}{50} \text{ x } (131.25 - 115)$

$= 4 + \frac{16.25}{50}$

$P_{75} = 4 + 0.325$

$P_{75} = 4.325$

अशा प्रकारे 75% फाइल्स परत येण्याचा कालावधी 4.325 महिने आहे

P_{60} = value of $=\frac{60(175)}{100}$ th observation

value of$=105\ th$ observation

P_{60} वर्गात 3-4आहे

Here L=3, h=1, f=40, c.f. =75 ,N=175

$P_{60} = L + \frac{h}{f} x (\frac{60N}{100} - c.f.)$

$= 3 + \frac{1}{40} x (\frac{60 X 175}{100} - 75)$

$= 3 + \frac{1}{40} x (105 - 75)$

$= 3 + \frac{30}{40}$

$= 3+0.75$

$P_{60} = 3.75$

अशा प्रकारे 60% फाइल्स परत येण्याचा कालावधी 3.75 महिने आहे

उदाहरण-1.15 विद्यार्थी विशिष्ट वर्गात गैरहजर राहिलेल्या दिवसांच्या वारंवारतेचे वितरण खालीलप्रमाणे आहे:

अनुपस्थित दिवसांची संख्या	विद्यार्थ्यांची संख्या
Below 10	15
Below 20	35
Below 30	60
Below 40	84
Below 50	106
Below 60	120
Below 70	125

उत्तर: गटांमध्ये डेटा रूपांतरित करा आणि संचयी फ्रेकेंव्सी सारणीपेक्षा कमी खालीलप्रमाणे तयार करा:

अनुपस्थित दिवसांची संख्या	विद्यार्थ्यांची संख्या	c.f पेक्षा कमी
0-10	15	15
10-20	20	35
20-30	25	60 (P_{30})
30-40	24	84 (D_6)
40-50	22	106
50-60	14	120
60-70	5	125

D_6 = value of $=\frac{6(125)}{10}$ observation

value of $=75th$ observation

D_6 वर्गात 30-40 आहे

Here L=30, h=10, f=24, c.f. =60 ,N=125

$D_6 = L + \frac{h}{f} x (\frac{6N}{10} - c.f.)$

$= 30 + \frac{10}{24} x (\frac{6 X 125}{10} - 60)$

$= 30 + \frac{10}{24} x (75 - 60)$

$= 30 + \frac{150}{24}$

= 30+ 6.25

D_6 = 36.25

म्हणजे 60% विद्यार्थी 36 दिवसांपेक्षा कमी दिवस गैरहजर असतात.

P_{30} = value of $=\frac{30(125)}{100}$ observation

P_{30} = value of $=37.5th$ observation

P_{30} वर्गात 20-30आहे

Here L=20, h=10, f=25, c.f. =35 ,N=125

$P_{30} = L + \frac{h}{f} x (\frac{30N}{100} - c.f.)$

$= 20 + \frac{10}{25} x (\frac{30 X 125}{100} - 35)$

$= 20 + \frac{10}{25} x (37.5 - 35)$

$= 20 + \frac{10}{25} x (2.5)$

= 20+ 1

P_{30} = 21

म्हणजे 30% विद्यार्थी 21 दिवसांपेक्षा कमी दिवस गैरहजर असतात.

सराव उदाहरण (Exercise)

1. खालील डेटासाठी D6 आणि P85 ची गणना करा:

79,82,36,38,51,72,68,70,64,63

2. पंधरा (15) मजुरांची दैनंदिन मजुरी खालीलप्रमाणे आहे:

230,400,350,200,250,380,210,225,375,180,375,450,300,350,250

D8 आणि P90 ची गणना करा

3. खालील साठी 2रा डेसील आणि 65वा पर्सेंटाइल मोजा:

x	80	100	120	145	200	280	310	380	400	410
f	15	18	25	27	40	25	19	16	8	7

4. खालील डेटावरून 15व्या, 65व्या आणि 91व्या घराच्या भाड्याची गणना करा.

घर भाडे (रु मध्ये)	11000	12000	13000	14000	15000	16000	17000	18000
घरांची संख्या	25	17	13	14	15	8	6	2

5. खालील फ्रेकेंव्सी वितरण वर्गातील विद्यार्थ्यांचे वजन दर्शवते.

वजन (किलोमध्ये)	40	45	50	55	60	65
विद्यार्थ्यांची संख्या	15	40	29	21	10	5

a) ज्या विद्यार्थ्यांचे वजन 50 किलोपेक्षा जास्त आहे त्यांची टक्केवारी शोधा.
b) प्रदान केलेला वजन स्तंभ मध्यभागी असल्यास मूल्ये नंतर विद्यार्थ्यांची टक्केवारी शोधा ज्यांचे वजन ५० किलोपेक्षा जास्त आहे.

6. खालील डेटावरून D4 आणि P48 ची गणना करा:

मीडियन मूल्य	2.5	7.5	12.5	17.5	22.5	Total
फ्रेकेंव्सी	7	18	25	30	20	100

7. खालील वितरणातील D9 आणि P20 ची गणना करा.

लांबी (इंच मध्ये)	0-20	20-40	40-60	60-80	80-100	100-120
युनिट्सची संख्या	1	14	35	85	90	15

1.5 Geometric Mean, Harmonic Mean and Combined Mean for given data

तुम्हाला दिलेल्या आकड्यांच्या संचासाठी भूमितीय मीडियन (Geometric Mean), हार्मोनिक मीडियन (Harmonic Mean) आणि संमिश्र मीडियन (Combined Mean) कसे काढायचे हे मराठीत समजावून सांगतो.

1. भूमितीय मीडियन (Geometric Mean): भूमितीय मीडियन हा एक प्रकारचा मीडियन (average) आहे जो गुणाकाराचा वापर करतो. ही पद्धत विशेषतः विविध गुणाकारांच्या संचासाठी उपयुक्त आहे.

सूत्र:

$$\text{G.M.} = \left(\prod_{i=1}^{n} xi\right)^{\frac{1}{n}}$$

उदाहरण:

समजा, आपल्याकडे आकड्यांचा संच आहे: 2, 8, 4

Step 1: सर्व संख्यांचा गुणाकार करा. 2×8×4=642 \times 8 \times 4 = 642×8×4=64

Step 2: त्या गुणाकाराचे n-व्या मूळ काढा (इथे n=आहे).

G.M.= $\sqrt[3]{64}$ = 4

2. हार्मोनिक मीडियन (Harmonic Mean): हार्मोनिक मीडियन हा एक प्रकारचा मीडियन आहे जो संख्यांचा व्युत्क्रम (reciprocals) वापरून काढला जातो.

सूत्र:

$$\text{H.M} = \frac{n}{\sum_{i=1}^{n} \frac{1}{xi}}$$

उदाहरण:

समजा, आपल्याकडे आकड्यांचा संच आहे: 2, 8, 4

Step 1: सर्व संख्यांचे व्युत्क्रम काढा आणि त्यांची बेरीज करा.

$\frac{1}{2} + \frac{1}{8} + \frac{1}{4}$ = 0.5 + 0.125 + 0.25 = 0.875

Step 2: संख्यांचा एकूण संच ~~nnn~~ (इथे n=3) वापरून हार्मोनिक मीडियन काढा.

H.M.= $\frac{3}{0.875}$ ≈ 3.43

3. संमिश्र मीडियन (Combined Mean): संमिश्र मीडियन काढताना दोन किंवा अधिक संचांचे एकत्रित मीडियन काढले जाते.

सूत्र:

$$\overline{X} = \frac{n_1\overline{X}_1 + n_2\overline{X}_2 + \ldots\ldots n_k X_k}{n_1 + n_2 + n_3 + \cdots + n_n}$$

इथे X¯i हे प्रत्येक संचाचे मीडियन आहे आणि ni हे त्या संचातील तत्वांची संख्या आहे.

उदाहरण:

समजा, आपल्याकडे दोन संच आहेत:

संच 1: 2, 8, 4 (मीडियन $\overline{X}_1$)

संच 2: 3, 7 (मीडियन$\overline{X}_2$)

Step 1: प्रत्येक2+8+4 संचाचे साधारण मीडियन काढा.

संच 1: X¯1= $\frac{2+8+4}{3}$ = $\frac{14}{3}$ ≈ 4.67

संच 2: X¯2= $\frac{3+7}{2}$ =$\frac{10}{2}$ =5

Step 2: दोन्ही संचांचे एकत्रित मीडियन काढा.

$\bar{X} = \frac{3\times4.67++2\times\ 5}{3+2} = \frac{24.01}{5} \approx 4.80$

सारांश:

- भूमितीय मीडियन (Geometric Mean): गुणाकाराच्या आधारावर काढलेला मीडियन.
- हार्मोनिक मीडियन (Harmonic Mean): व्युत्क्रमांच्या आधारे काढलेला मीडियन.
- संमिश्र मीडियन (Combined Mean): दोन किंवा अधिक संचांचे एकत्रित साधारण मीडियन.

हे तिन्ही मीडियन विविध प्रकारच्या आकड्यांच्या संचासाठी उपयुक्त आहेत आणि आपल्या गरजेनुसार त्यांचा वापर करता येतो.

1.6 मोड शोधण्यासाठी ग्राफिकल प्रतिनिधित्व (हिस्टोग्राम आणि मध्य ओगिव्ह वक्र)

हिस्टोग्राम वापरून ग्राफिक पद्धतीने डेटा सेटचा मोड शोधण्यासाठी, तुम्ही मोड दर्शविण्यासाठी हिस्टोग्रामच्या सर्वात उंच शिखराचा वापर करू शकता. मोड हे डेटा सेटमध्ये वारंवार दिसणारे मूल्य आहे.

गटबद्ध डेटासाठी, मोड हा मोडल क्लासचा मध्यबिंदू आहे. सर्वोच्च बार वापरून मोडचा अंदाज घेण्यासाठी, तुम्ही मोड बहुभुज तयार करू शकता किंवा इतर इंटरपोलेशन तंत्र वापरू शकता. तुम्ही सर्वात उंच पट्टीच्या वरच्या उजव्या कोपऱ्यापासून बारच्या वरच्या उजव्या कोपऱ्यापर्यंत एक रेषा काढू शकता जी वर्गाची फ्रेकेंव्सी दर्शवते. नंतर, या रेषांच्या छेदनबिंदूपासून x-अक्षापर्यंत एक उभी रेषा काढा. x-अक्षावरील मूल्य मोडसाठी अंदाज दर्शवते.

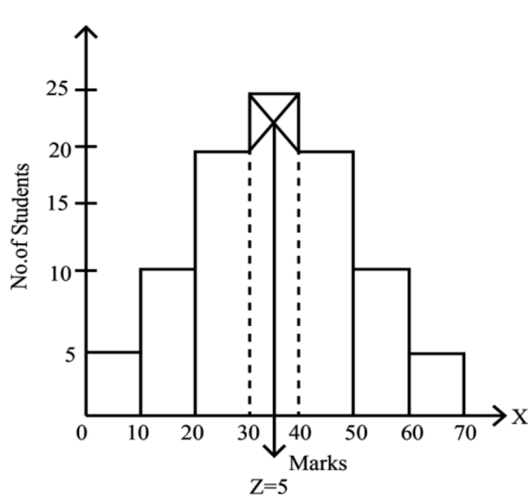
ग्राफिकल पद्धतीनुसार मोड

1. दिलेल्या मालिकेच्या संचाचा हिस्टोग्राम काढा.
2. सर्वात जास्त उंचीसह हिस्टोग्रामचा आयत मालिकेचा मॉडेल वर्ग असेल.
3. मोडल वर्गाच्या आयताच्या वरच्या डाव्या कोपऱ्यात/बिंदूला जोडणारी एक रेषा काढा.

उदाहरण-1.16 ग्राफिक पद्धतीने मोड शोधा:

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No of Student	5	10	20	25	20	20	5

उत्तर: प्रथम, आम्ही दिलेल्या डेटाचा हिस्टोग्राम काढतो:



हिस्टोग्रामवरून हे स्पष्ट होते की मोडचे मूल्य 35 आहे.

समान प्रश्न

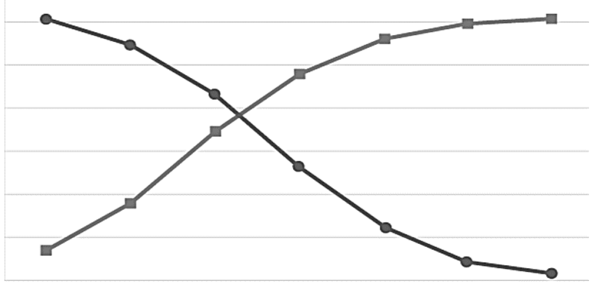
1.खालील डेटावरून मोडची गणना करा:

वय	0-5	5-10	10-15	15-20	20-25	25-30	30-35
रुग्णांची संख्या	6	11	18	24	17	13	5

2. खालील डेटावरून मोडची गणना करा

गुण	10-20	20-30	30-40	40-50	50-60	60-70
विद्यार्थ्यांची संख्या	8	11	9	25	12	16

- **Ogive** हा शब्द आर्किटेक्चरमध्ये वक्र किंवा वक्र आकारांचे वर्णन करण्यासाठी वापरला जातो. Ogives हे आलेख आहेत जे डेटामधील विशिष्ट चल किंवा मूल्याच्या खाली किंवा वर किती संख्या आहेत याचा अंदाज लावण्यासाठी वापरला जातो. ऑगिव्ह तयार करण्यासाठी, प्रथम, व्हेरिएबल्सची एकत्रित फ्रेकेंव्सी फ्रेकेंव्सी सारणी वापरून मोजली जाते. दिलेल्या डेटा सेटमध्ये मागील सर्व व्हेरिएबल्सची फ्रिक्वेन्सी जोडून हे केले जाते. संचयी फ्रेकेंव्सी सारणीतील परिणाम किंवा शेवटची संख्या नेहमी चलांच्या एकूण फ्रिक्वेन्सीच्या समान असते. फ्रेकेंव्सी वितरणाचे सर्वात सामान्यपणे वापरले जाणारे आलेख म्हणजे हिस्टोग्राम, फ्रेकेंव्सी बहुभुज, फ्रेकेंव्सी वक्र आणि ओगिव्स (संचयी फ्रेकेंव्सी वक्र). "Ogive" नावाच्या एका आलेखाची सविस्तर चर्चा करूया.
- **Ogive व्याख्या:** Ogive ची व्याख्या मालिकेतील फ्रेकेंव्सी वितरण आलेख म्हणून केली जाते. Ogive हा संचयी वितरणाचा आलेख आहे, जो क्षैतिज समतल अक्षावरील डेटा मूल्ये आणि एकतर संचयी सापेक्ष फ्रिक्वेन्सी, संचयी फ्रिक्वेन्सी किंवा उभ्या अक्षावरील संचयी टक्के फ्रिक्वेन्सी स्पष्ट करतो. संचयी फ्रेकेंव्सी वर्तमान बिंदूपर्यंत सर्व मागील फ्रिक्वेन्सीची बेरीज म्हणून परिभाषित केली जाते. प्रत्येक वर्ग अंतराच्या संचयी वारंवारतेशी संबंधित बिंदू प्लॉट करून ऑगिव्ह तयार करा. बहुतेक सांख्यिकीशास्त्रज्ञ सचित्र प्रतिनिधित्वातील डेटा स्पष्ट करण्यासाठी ओगिव्ह वक्र वापरतात. हे विशिष्ट मूल्यापेक्षा कमी किंवा समान असलेल्या निरीक्षणांच्या संख्येचा अंदाज लावण्यास मदत करते.
- **Ogive आलेख:** फ्रेकेंव्सी वितरणाचे आलेख हे फ्रेकेंव्सी आलेख आहेत जे स्वतंत्र आणि सतत डेटाची वैशिष्ट्ये प्रदर्शित करण्यासाठी वापरले जातात. सारणीबद्ध डेटापेक्षा असे आकडे डोळ्यांना अधिक आकर्षित करतात. हे आम्हाला दोन किंवा अधिक फ्रेकेंव्सी वितरणाचा तुलनात्मक अभ्यास करण्यास मदत करते. आपण दोन फ्रेकेंव्सी वितरणाचा आकार आणि नमुना संबंधित करू शकतो. Ogives च्या दोन पद्धती आहेत:
  1. Ogive पेक्षा कमी
  2. Ogive पेक्षा जास्त
  3. Ogive वक्र पेक्षा कमी आणि जास्त



वर दिलेला आलेख Ogive वक्र पेक्षा कमी आणि जास्त दर्शवतो. उगवणारा वक्र (गुलाबी वक्र) ओगिव्हपेक्षा कमी दर्शवितो आणि घसरणारा वक्र (जांभळा वक्र) ओगिव्हपेक्षा मोठा दर्शवितो.

- **Ogive पेक्षा कमी:** सर्व आधीच्या वर्गांची फ्रेकेंव्सी वर्गाच्या वारंवारतेमध्ये जोडली जाते. या मालिकेला संचयीपेक्षा कमी मालिका म्हणतात. प्रथम श्रेणी फ्रेकेंव्सी द्वितीय श्रेणी फ्रेकेंव्सी आणि नंतर तृतीय श्रेणी फ्रेकेंव्सी जोडून ते तयार केले जाते.
- **Ogive पेक्षा जास्त:** त्यानंतरच्या वर्गांची फ्रेकेंव्सी वर्गाच्या वारंवारतेमध्ये जोडली जाते. या मालिकेला संचयी मालिकेपेक्षा जास्त म्हणतात. एकूण मधून प्रथम श्रेणी, द्वितीय श्रेणी फ्रेकेंव्सी वजा करून, तृतीय श्रेणी फ्रेकेंव्सी वजा करून ते तयार केले जाते. ऊर्ध्वगामी संचय परिणाम संचयी मालिकेपेक्षा जास्त आहे.
- **Ogive चार्ट:** ऑगिव्ह चार्ट हा संचयी फ्रेकेंव्सी वितरण किंवा संचयी सापेक्ष फ्रेकेंव्सी वितरणाचा वक्र आहे. असा वक्र काढण्यासाठी, फ्रिक्वेन्सी एकूण वारंवारतेच्या टक्केवारीच्या रूपात व्यक्त केल्या पाहिजेत. मग, अशा टक्केवारी एकत्रित केल्या जातात आणि प्लॉट केल्या जातात, जसे ओगिव्हच्या बाबतीत. खाली Ogive पेक्षा कमी आणि मोठे बांधण्यासाठी पायऱ्या आहेत.

Ogive वक्र पेक्षा कमी कसे काढायचे?

- क्षैतिज आणि अनुलंब अक्ष काढा आणि चिन्हांकित करा.
- y-अक्ष (उभ्या अक्ष) आणि x-अक्षावर (क्षैतिज अक्ष) वरच्या श्रेणीच्या मर्यादांसह एकत्रित फ्रिक्वेन्सी घ्या.
- प्रत्येक उच्च-श्रेणी मर्यादेच्या विरूद्ध, संचयी फ्रिक्वेन्सी प्लॉट करा.
- सतत वक्र सह बिंदू कनेक्ट करा.

Ogive वक्र पेक्षा मोठे किंवा जास्त कसे काढायचे?

- क्षैतिज आणि अनुलंब अक्ष काढा आणि चिन्हांकित करा.
- y-अक्ष (उभ्या अक्ष) बाजूने संचयी फ्रिक्वेन्सी आणि x-अक्ष (क्षैतिज अक्ष) वरील निम्न-वर्ग मर्यादा घ्या.
- प्रत्येक निम्न-वर्गाच्या मर्यादेच्या विरूद्ध, संचयी फ्रिक्वेन्सी प्लॉट करा.
- सतत वक्र सह बिंदू कनेक्ट करा.

ओगिव्ह कर्व्हचा उपयोग

दिलेल्या डेटाच्या संचाचा मध्य शोधण्यासाठी ऑगिव्ह आलेख किंवा संचयी फ्रेकेंव्सी आलेख वापरले जातात. एकाच आलेखावर एकत्रित फ्रेकेंव्सी वक्र, पेक्षा कमी आणि पेक्षा मोठे दोन्ही काढले असल्यास, आपण मीडियनमूल्य सहज शोधू शकतो. ज्या बिंदूमध्ये, दोन्ही वक्र, x-अक्षाशी संबंधित, छेदतात, तो मीडियनमूल्य देतो. मीडियनशोधण्याव्यतिरिक्त, Ogives डेटा सेट मूल्यांच्या टक्केवारीची गणना करण्यासाठी वापरले जातात.

Ogive उदाहरण

उदाहरण-1.17: संचयी फ्रेकेंव्सी सारणीपेक्षा अधिक तयार करा आणि खाली दिलेल्या डेटासाठी ऑगिव्ह काढा.

गुण	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80
फ्रेकेंव्सी	3	8	12	14	10	6	5	2

उत्तर: संचयी फ्रेकेंव्सी सारणी "पेक्षा जास्त":

गुण	फ्रेकेंव्सी	संचयी फ्रेकेंव्सी पेक्षा जास्त
पेक्षा जास्त 1	3	60
पेक्षा जास्त 11	8	57
पेक्षा जास्त 21	12	49
पेक्षा जास्त 31	14	37
पेक्षा जास्त 41	10	23
पेक्षा जास्त 51	6	13
पेक्षा जास्त 61	5	7
पेक्षा जास्त 71	2	2

एक ओगिव्ह प्लॉटिंग:

(70.5, 2), (60.5, 7), (50.5, 13), (40.5, 23), (30.5, 37), (20.5, 49), (10.5, 57), (0.5, 60).

ऑगिव्ह हे x-अक्षावरील एका बिंदूशी जोडलेले असते, जे शेवटच्या वर्गाची वास्तविक वरची मर्यादा दर्शवते, म्हणजे (80.5, 0)

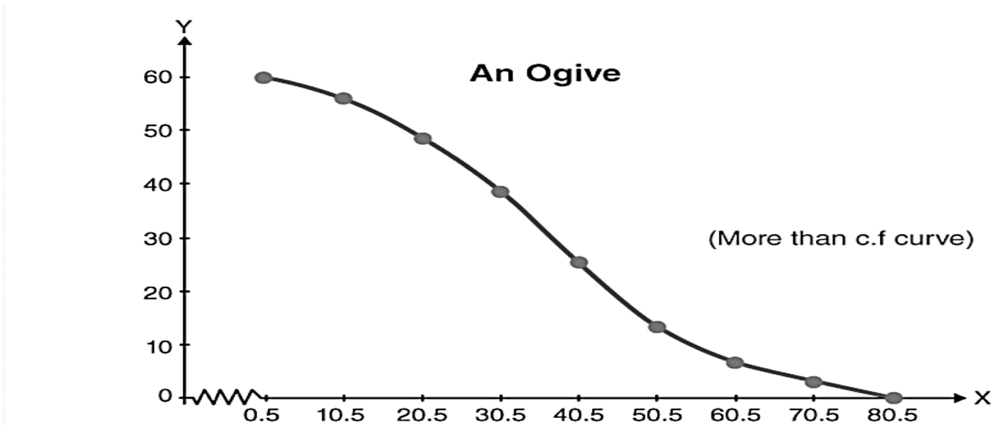
x-अक्ष, 1 सेमी = 10 गुण घ्या

Y-अक्ष = 1 सेमी - 10 c.f

Ogive वक्र पेक्षा अधिक:

Y-axis = 1 cm – 10 c.f

More than the Ogive Curve:



सराव प्रश्न:

खालील फ्रिक्वेन्सी सारणीसाठी ऑगिव्ह पेक्षा अधिक तयार करा

Lass Interval	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequency	5	15	20	23	17	11	9

1.7 Measure of dispersion

श्रेणी (Range): 'श्रेणी' हे टोकाच्या मूल्यांमधील फरक दर्शवते. दोन्ही टोकांकडील मूल्यांमधील परिवर्तनशीलता मोजण्यासाठी (म्हणजे सर्वात लहान आणि सर्वात मोठी मूल्ये) वापरतो.

- **गट न केलेल्या डेटासाठी श्रेणी (The Range for Ungrouped data)**

श्रेणी हा सर्वात मोठा मूल्ये आणि सर्वात लहान मूल्ये मधील फरक आहे. तो 'R' ने दर्शविला जातो.

श्रेणी = सर्वात मोठे डेटा मूल्य - सर्वात लहान डेटा मूल्य.

L = सर्वात मोठे मूल्य आणि S = सर्वात लहान मूल्य, नंतर

R = L-S

- **गटबद्ध डेटासाठी श्रेणी (The Range for Ungrouped data):** फ्रेकेंव्सी वितरणामध्ये (गटबद्ध किंवा सतत फ्रेकेंव्सी), श्रेणी ही वर्गाची वरच्या टोकावरील मर्यादा आणि वितरणाच्या खालच्या टोकावरील वर्गाची खालची मर्यादा यांच्यातील फरक मानली जाते.

  श्रेणी (सर्वोच्च वर्गाची सर्वात मोठी मर्यादा) – (सर्वात खालच्या वर्गाची सर्वात कमी मर्यादा)

- **श्रेणीचे गुणांक (Coefficient of range)**

श्रेणीचे गुणांक असे परिभाषित केले आहे,

श्रेणीचे गुणांक = $\frac{L-S}{L+S}$

Coefficient of range= $\frac{L-S}{L+S}$

खालील वितरणाची गुणांक = $\frac{\text{सर्वात मोठे मूल्य} - \text{सर्वात लहान मूल्य}}{\text{सर्वात मोठे मूल्य} + \text{सर्वात लहान मूल्य}}$

उदाहरण-1.18: खालील वितरणाची श्रेणी शोधा

2, 3, 1, 10, 6, 31, 17, 20, 24

उत्तर: आम्हाला माहीत आहे, Range= L-S

जेथे L सर्वात मोठे मूल्य S=सर्वात लहान मूल्य दिले आहे,

2, 3, 1, 10, 6, 31, 17, 20, 24 या मूल्यांमधून,

सर्वात मोठे मूल्य L= 31 सर्वात लहान मूल्य = S = 1

Range= 31-1

Range = 30

उदाहरण-1.19: खालील वितरणाची श्रेणी शोधा:

42, 39, 40, 48, 41, 45, 44

उत्तर: आम्हाला माहित आहे, Range = L-S

L= सर्वात मोठे मूल्य S सर्वात लहान मूल्य दिलेली मूल्ये आहेत,

39, 40, 41, 42, 44, 45, 45, 48

निरीक्षण करा की, या मूल्यांमधून,

सर्वात मोठे मूल्य L=48 सर्वात लहान मूल्य = S=39

Range = 48-39

Range = 9

उदाहरण-1.20: खालील वितरणाची श्रेणी शोधा:

3, 6, 10, 1, 15, 16, 21, 19, 18

उत्तर: आम्हाला माहित आहे Range =L-S

L= वितरणाचे सर्वात मोठे मूल्य S= वितरणाचे सर्वात लहान मूल्य

दिलेले वितरण, 3, 6, 10, 1, 15, 16, 21, 19, 18 पाहा की,

या मूल्यांमधून, सर्वात मोठे मूल्य = L=21 सर्वात लहान मूल्य S=1

Range = 21-1

Range= 20

उदाहरण-1.21: खालील डेटावरून श्रेणीची गणना करा:

किलोमध्ये वजन: 70, 75, 69, 80, 85, 83, 65, 89, 73, 84, 90,

उत्तर: आम्हाला माहिती आहे. Range = L-S

70, 75, 69, 80, 85, 83, 65, 8 73,84,90 या मूल्यांमधून,

सर्वात मोठे मूल्य L=90   सर्वात लहान मूल्य = S=65

Range = 90-65

Range = 25

उदाहरण-1.22: खालील डेटाची श्रेणी शोधा:

800, 725, 750, 900, 925, 910, 1000, 790, 870, 920

उत्तर: दिलेली, श्रेणी 800, 725, 750, 900, 925, 910, 700, 700, 700, 920

Range= सर्वात मोठे मूल्य-सर्वात लहान मूल्य

= 1000-725

= 275

उदाहरण-1.23: शोधा (i) श्रेणी (ii) श्रेणीचे गुणांक खालील डेटाच्या श्रेणीचे:

50, 90, 120, 40, 180, 200, 80.

उत्तर: आम्हाला माहित आहे, Range = L-S

आणि श्रेणीचे गुणांक = $\frac{L-S}{L+S}$

Coefficient of range= $\frac{L-S}{L+S}$

L= डेटाचे सर्वात मोठे मूल्य)

S दिलेले डेटाचे सर्वात लहान मूल्य,

50, 90, 120, 40, 180, 200, 80 या मूल्यांमधून,

सर्वात मोठे मूल्य = L = 200

सर्वात लहान मूल्य = S = 40

(i) श्रेणी

Range = L-S

= 200-40

Range = 160 वरून

(ii) श्रेणीचे गुणांक

Coefficient of range= $\frac{L-S}{L+S}$

$= \frac{200-40}{200+40}$

$= \frac{160}{240}$

$= \frac{2}{3}$

उदाहरण-1.23: खालील वितरणाच्या श्रेणीची श्रेणी आणि गुणांक शोधा.

Xi	10	20	30	40	50
fi	7	5	3	2	1

उत्तर: आम्हाला माहित आहे, Range = L-S

आणि श्रेणीचे गुणांक = $\frac{L-S}{L+S}$
L= डेटाचे सर्वात मोठे मूल्य
S दिलेले डेटाचे सर्वात लहान मूल्य,
दिलेले वितरण,

Xi	10	20	30	40	50
fi	7	5	3	2	1

या तक्त्यावरून ते पहा
L = 50 चे सर्वात मोठे मूल्य
S = 10 चे सर्वात लहान मूल्य
(i) श्रेणी
Range = L-S
= 50-10
Range = 40
(ii) श्रेणीचे गुणांक= $\frac{L-S}{L+S}$
$= \frac{50-10}{50+10}$
$= \frac{40}{60}$
$= \frac{2}{3}$

सराव (Exercise)
1. खालील डेटाची श्रेणी शोधा : 10, 5, 12, 2, 15, 20, 8, 10
2.खालील वितरणाच्या श्रेणीचा (i) श्रेणी (ii) गुणांक शोधा

Marks	0-10	10-20	20-30	30-40	40-50
No. of students	8	12	10	15	5

3. खालील डेटावरून, गुणांक आणि श्रेणीची श्रेणी काढा

Marks	10-19	20-29	30-39	40-49	50-59	60-69
No. of students	6	10	16	14	8	4

सरासरी विचलन (The Mean Deviation): जर n ची X1, X2, X3,......, X, X चे अंकगणित Arithmatic mean $\bar{x}$ असेल, तरx_1 - $\bar{x}$, x_2 - $\bar{x}$, x_3 - $\bar{x}$, x-$\bar{x}$ यांना X1, X2, X3, अंकगणित मध्यापासूनची deviation म्हणतात.

$$\sum_{i=1}^{n} \mathrm{xi} - \bar{x} = \sum_{i=1}^{n} \mathrm{xi} - \sum_{i=1}^{n} \bar{x} = 0$$

$$\Sigma(x-x) = \Sigma X_1 - \sum x = nx-nx=0$$

सरासरी विचलन (किंवा, अधिक अचूकपणे, सरासरी ओलुट विचलन) सूत्राद्वारे परिभाषित केले जाते.

M.D. = $\frac{\sum_{i=1}^{n}|\mathrm{xi} - \bar{x}|}{n}$
$= \frac{\sum|di|}{N}$

Where,

- $|di|$=|xi $-$ $\overline{x}$|
- $\overline{x}$ = Arithmetic mean

The Mean Deviation for Row data

मीन बद्दल सरासरी विचलन(Mean Deviation about mean)

$$=\frac{\sum|\text{xi} - \overline{x}|}{N}$$

- $\overline{x}$= mean of N observation

मीडियन बद्दल सरासरी विचलन(Mean Deviation about median)

$$=\frac{\sum|\text{xi} - \text{M}|}{N}$$

- M = median of N observations

डिस्क्रिट फ्रिक्वेन्सी डिस्ट्रिब्युशन साठी सरासरी विचलन Mean Deviation for Discrete Frequency Distribution

Mean Deviation about mean

$$=\frac{\sum fi|\text{xi} - \overline{x}|}{N} = \frac{\sum fi|\text{di}|}{N}$$

$\overline{x}$= mean of N observation

मीडियन बद्दल सरासरी विचलन(Mean Deviation about median)

$$=\frac{\sum fi|\text{xi} - \text{M}|}{N} = \frac{\sum fi|\text{di}|}{N}$$

M = median of N observations

Mean Deviation for grouped Frequency Distribution

मीन बद्दल सरासरी विचलन(Mean Deviation about mean)

$$=\frac{\sum fi|\text{xi} - \overline{x}|}{N} = \frac{\sum fi|\text{di}|}{N}$$

$\overline{x}$= mid value or class mark

मीडियन बद्दल सरासरी विचलन(Mean Deviation about median)

$$=\frac{\sum fi|\text{xi} - \text{M}|}{N} = \frac{\sum fi|\text{di}|}{N}$$

M = median of Distribution

उदाहरण-1.24: खालील डेटासाठी मध्यकापासून सरासरी विचलन शोधा:

Class-interval	0-10	10-20	20-30	30-40	40-50
Frequency	5	8	15	16	6

उत्तर: आम्हाला माहिती आहे,

मीडियन बद्दल सरासरी विचलन(Mean Deviation about median)

Class-interval	**Middle value(xi)**	**Frequency(fi)**	**c.f.**	$|di|$=|xi $-$ $\overline{x}$|	**fi**$|di|$
0-10	5	5	5	20	100
10-20	15	8	13	10	80
20-30	25	15	28	0	0
30-40	35	16	44	10	160

40-50	45	6	50	20	120
		N=50			$\sum fi\|di\|$ =460

Median=M=$(\frac{50+1}{2})$ th observation=$\frac{51}{2}$= 25.5 it lies between class 20-30

Median=M=25

$=\frac{\sum fi|xi-\bar{x}|}{N}=\frac{\sum fi|di|}{N}$

$=\frac{460}{50}= 9.2$

मीडियन बद्दल सरासरी विचलन =9.2

उदाहरण-1.25: खालील डेटासाठी सरासरी विचलनाची गणना करा

Marks	0-10	10-20	20-30	30-40	40-50
No of students	5	8	15	16	6

उत्तर: आम्हाला माहिती आहे,

$=\frac{\sum fi|xi-\bar{x}|}{N}=\frac{\sum fi|di|}{N}$

$\bar{x}$= mean of N observation

$\bar{x}$= mean $=\frac{\sum fixi}{N}$

Class-interval	Middle value(xi)	Frequency(fi)	fixi	$\|di\|$=$\|xi - \bar{x}\|$	fi$\|di\|$
0-10	5	5	25	22	110
10-20	15	8	120	12	96
20-30	25	15	375	2	30
30-40	35	16	560	8	128
40-50	45	6	270	18	108
		N=50	$\sum fixi = 1350$		$\sum fi\|di\|$ =472

$\bar{x}$= mean $=\frac{\sum fixi}{N}$

$\bar{x}=\frac{1350}{50}= 27$

मीन बद्दल सरासरी विचलन (Mean Deviation about mean)

$=\frac{\sum fi|di|}{N}$

$=\frac{472}{50}$

= 9.44

मीन बद्दल सरासरी विचलन(Mean Deviation about mean)=9.44

उदाहरण-1.25: खालील डेटासाठी सरासरी विचलनाची गणना करा

xi	3	4	5	6	7	8
fi	4	9	10	8	6	3

उत्तर: आम्हाला माहिती आहे,

$\bar{x} = \frac{\sum fi|xi - \bar{x}|}{N} = \frac{\sum fi|di|}{N}$

$\bar{x}$= mean of N observation

$\bar{x}$= mean $= \frac{\sum fixi}{N}$

xi	fi	fixi	$|di|$=$|xi - \bar{x}|$	fi$|di|$
3	4	12	2.3	9.2
4	9	36	1.3	11.2
5	10	50	0.3	3
6	8	48	0.7	5.6
7	6	42	1.7	10.2
8	3	24	2.7	8.1
	N=40	$\sum fixi = 212$		$\sum fi|di|$ =47.8

$\bar{x}$= mean $= \frac{\sum fixi}{N}$

$\bar{x} = \frac{212}{40} = 5.3$

मीन बद्दल सरासरी विचलन(Mean Deviation about mean)

$= \frac{\sum fi|di|}{N}$

$= \frac{47.8}{40}$

=1.195

मीन बद्दल सरासरी विचलन (Mean Deviation about mean) nearby 1.20

मानक विचलन- स्टँडर्ड डेव्हिएशन (S.D.) Standard Deviation: मानक विचलन हे सर्व डेटा मूल्यांवर अवलंबून असण्याचे एक माप आहे. प्रत्येक डेटा मूल्य आणि सरासरीमधील एकूण फरक शोधून त्याची गणना केली जाते. जर $\bar{x}$ हा डेटाच्या संचाचा सरासरी असेल तर परिमाण $xi - \bar{x}$ याला मध्यापासूनचे विचलन म्हणतात. दिलेल्या मूल्यांच्या त्यांच्या अंकगणितीय मध्यापासून विचलनाच्या वर्गांच्या अंकगणित सरासरीचे धनात्मक वर्गमूळ म्हणून त्याची व्याख्या केली जाते.हे 'σ' या चिन्हाने दर्शविले जाते.

$|di|=|xi - \bar{x}|$

and $\sum fi = N$

$$\bar{x} = \text{mean} = \frac{\sum fixi}{N}$$

$$\text{Coefficient of Standard deviation} = \frac{S.D.}{Mean} = \frac{\sigma}{\bar{x}}$$

भिन्नता Variance: मानक विचलनाच्या (S.D.) वर्गाला प्रसरण(Variance) असे म्हणतात आणि ते "σ^2" ने दर्शविले जाते. भिन्नता हे परिवर्तनशीलतेचे एक माप आहे जे सर्व डेटा वापरते. प्रसरण हा प्रत्येक डेटा मूल्य आणि सरासरीमधील वर्ग फरकांची प्रमाणित विचलन सरासरी आहे.

मानक विचलन =$\sqrt{\text{भिन्नता}}$

$$\text{Standard deviation} = \sqrt{variance}$$

Variance = $(S.D.)^2$ किंवा σ^2

डेटा सेटचे मानक विचलन हे विचरणाचे धनात्मक वर्गमूळ आहे. हे डेटाच्या समान युनिट्समध्ये मोजले जाते, ज्यामुळे ते भिन्नतेपेक्षा अधिक सहजपणे स्पष्ट होते.

$\sigma^2 = \frac{\sum fixi^2}{N} - \bar{x}^2$

σ = S.D. = $\sqrt{variance}$

Coefficient of Variance भिन्नतेचे गुणांक= $\frac{S.D.}{Mean}$ $X\ 100 = \frac{\sigma}{\bar{x}}$ X 100

उदाहरण-1.26: खालील डेटासाठी मानक विचलनाची गणना करा

1,2,3,4,5,6,7,8,9

उत्तर:

σ = S.D. = $\sqrt{\frac{\sum di^2}{N}}$ where di= xi $- \bar{x}$

$\bar{x} = \frac{\sum xi}{N}$

Xi	$\lvert di \rvert$=$\lvert$xi $- \bar{x}\rvert$	di^2
1	4	16
2	3	9
3	2	4
4	1	1
5	0	0
6	1	1
7	2	4
8	3	9
9	4	16
$\sum xi = 45$		$\sum di^2 = 60$

$\bar{x} = \frac{\sum xi}{N}$

$\bar{x} = \frac{45}{9}$

$\bar{x}$=5

Now S.D. = $\sqrt{\frac{60}{9}}$

$= \sqrt{\frac{20}{3}}$

σ = 2.58

σ = S.D. =2.58

उदाहरण-1.27: खालील डेटासाठी मानक विचलनाची गणना करा

साप्ताहिक खर्च रु. च्या खाली	5	10	15	25
विद्यार्थ्यांची संख्या	6	16	20	46

उत्तर:

$\bar{x}$= mean $= \frac{\sum fixi}{N}$

दिलेल्या डेटावरून सारणी मिळवा:

साप्ताहिक खर्च रु. च्या खाली(xi)	**विद्यार्थ्यांची संख्या (fi)**	**$f_i\ x_i$**	$\lvert di \rvert$=$\lvert$xi $- \bar{x}\rvert$ $\bar{x} = 18.64$	$d_i^{\ 2}$	fi$d_i^{\ 2}$
05	06	30	13.64	186.05	1116.3
10	16	160	8.64	74.65	1194.4

साप्ताहिक खर्च रु. च्या खाली(xi)	विद्यार्थ्यांची संख्या (fi)	$f_i x_i$	$\lvert di \rvert = \lvert xi - \bar{x} \rvert$ $\bar{x} = 18.64$	d_i^2	fid_i^2
15	20	300	3.64	13.25	265
25	46	1150	6.36	40.45	1860.7
	$\sum fi = N$=88	$\sum fi$xi=1640			$\sum fidi^2 = 4436.4$

$\bar{x}$= mean $= \frac{\sum fixi}{N} = \frac{1640}{88}$ =18.64

σ = S.D. = $\sqrt{\frac{\sum fidi^2}{N}}$

$= \sqrt{\frac{4436.4}{88}}$

$= \sqrt{50.4136}$

σ = 7.1

S.D. =7.1

उदाहरण-1.28: खालील डेटासाठी मानक विचलनाची गणना करा

वर्ग मध्यांतर	0-10	10-20	20-30	30-40	40-50
फ्रेकेंव्सी	3	5	8	3	1

उत्तर:

σ = S.D. = $\sqrt{\frac{\sum di^2}{N}}$ where di= xi $- \bar{x}$

$\bar{x} = \frac{\sum xi}{N}$

दिलेल्या डेटावरून सारणी मिळवा:

वर्ग मध्यांतर	मीडियन मूल्य (xi)	फ्रेकेंव्सी (fi)	fixi	$\lvert di \rvert = \lvert xi - \bar{x} \rvert$	d_i^2	fid_i^2
0-10	5	3	15	17	289	867
10-20	15	5	75	7	49	245
20-30	25	8	200	3	9	72
30-40	35	3	105	13	169	507
40-50	45	1	45	23	529	529
		N=20	$\sum fi$xi = 440			$\sum fidi^2 = 2220$

$\bar{x}$= mean $= \frac{\sum fixi}{N} = \frac{440}{20}$ =22

σ = S.D. = $\sqrt{\frac{\sum fidi^2}{N}}$

$=\sqrt{\frac{2220}{20}}$

$=\sqrt{111}$

σ = 10.54

S.D. =10.54

सराव-Practice Questions

प्रश्न 1. खालील डेटामधून सरासरी विचलन आणि सरासरी विचलनाच्या सह-कार्यक्षमतेची गणना करा:

Marks of the students: 86, 25, 87, 65, 58, 45, 12, 71, 35.

प्रश्न 2. खालील डेटासाठी सरासरी विचलनाची गणना करा.

X	12	9	6	18	10
f	7	3	8	1	2

प्रश्न 3. खालील डेटासाठी सरासरी विचलनाची गणना करा

वर्ग मध्यांतर	0-2	2-4	4-6	6-8
फ्रेकेंव्सी	4	2	5	3

प्रश्न 4. खालील डेटासाठी सरासरी, भिन्नता आणि मानक विचलनाची गणना करा

वर्ग मध्यांतर	0-10	10-20	20-30	30-40	40-50	50-60
फ्रेकेंव्सी	27	10	7	5	4	2

प्रश्न 5. खालील मूल्यांच्या मानक विचलनाची गणना करा:

5, 10, 25, 30, 50

प्रश्न 6. खालील डेटासाठी सरासरी आणि मानक विचलन शोधा

X	60	61	62	63	64	65	66	67	68
f	2	1	12	29	25	12	10	4	5

प्रश्न 7. डिझाइनमध्ये काढलेल्या वर्तुळांचे व्यास (मिमीमध्ये) खाली दिले आहेत:

व्यास	33-36	37-40	41-44	45-48	49-52
वर्तुळची संख्या	15	17	21	22	25

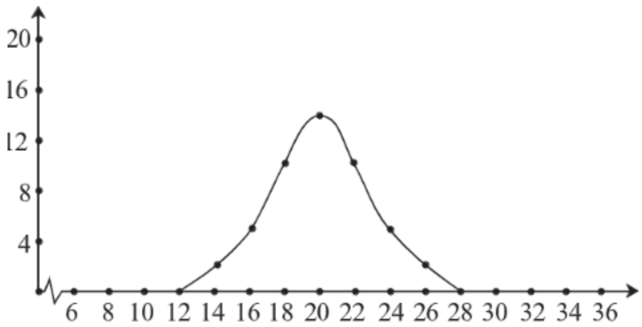
मानक विचलनाची गणना करा

1.8 तिरकसपणा- स्केवणेस (Skewness)

- **तिरकसपणाचे उपाय:** मध्यवर्ती प्रवृत्ती आणि फैलाव यांचे समान माप असलेली दोन वितरणे सममितीच्या दृष्टीने भिन्न असू शकतात. हे अपरिहार्यपणे महत्वाचे आहे कारण ते संख्यात्मक दृष्टीने डेटाचे दृश्यमान अर्थ देते. तसेच या संख्यांमुळे आम्हाला हे सूचित करण्यात मदत होते की वितरणामध्ये जास्त फ्रिक्वेन्सी (निरीक्षणांशी संबंधित) डावीकडे किंवा अक्षाच्या उजव्या टोकाकडे जमा होतात; हे सममितीचे एक माप आहे ज्याला Skewness म्हणतात.
- **तिरकसपणा (Skewness):** तिरकसपणा हे सूचित करते की वितरण सममितीय किंवा असममित आहे आणि असममित असल्यास, ते दिशासह असममितीची व्याप्ती देखील दर्शवते.
- **झिरो स्क्युनेस:** उदाहरणार्थ, असे फ्रिक्वेन्सी डिस्ट्रिब्युशन असते की, सरासरीपासून समान अंतरावर असलेल्या व्हेरिएबल्सची व्हॅल्यू समान फ्रिक्वेन्सी असतात. या प्रकरणात, वितरण सममितीय (शून्य स्क्युनेस) असल्याचे म्हटले जाते

X(चल)	14	16	18	20	22	24	26
f(फ्रेकेंव्सी)	2	5	10	14	10	5	2

वरील टेबल मध्ये मीन 20 आहे. लक्षात घ्या की x ची मूल्ये सरासरीपासून समदुष्टी आहेत.
समान फ्रिक्वेन्सी आहेत. उदा. 18 आणि 22, 20 (मध्य) पासून 2 च्या अंतरावर आहेत आणि अनुक्रमे 10 आणि 10 च्या समान फ्रेकेंव्सी आहेत.
हे सममितीय वितरणाचे संकेत आहे.

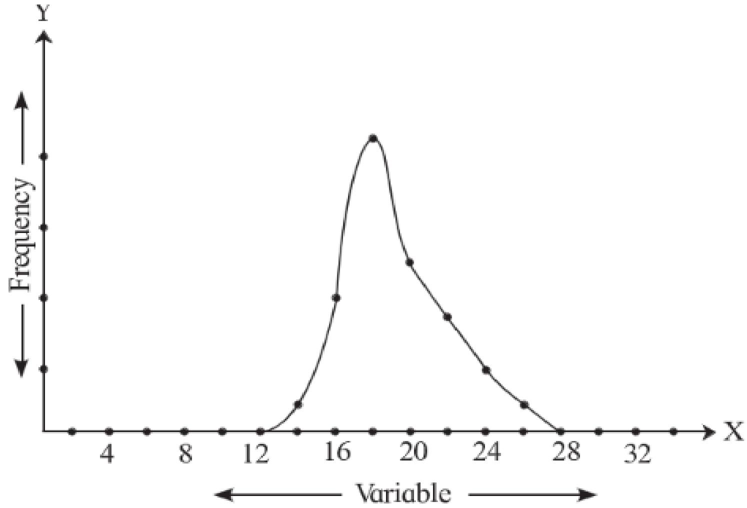


सममितीय वितरणाला मीन = मीडियन= मोड आहे
उदाहरणार्थ रॉडची लांबी, व्यक्तींचे वजन साधारणतः सममितीय असते.

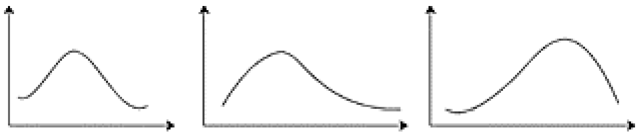
- **असममित वितरण (सकारात्मक विकृती):(Positive Skewness):** ज्या प्रकरणांमध्ये x ची मूल्ये जी मध्यापासून समान अंतरावर असतात त्यांची फ्रेकेंव्सी असमान असते, वितरण असममित असल्याचे म्हटले जाते. असममित वितरणास सकारात्मक तिरकस म्हटले जाते जर व्हेरिएबलच्या व्हॅल्यूजमध्ये मीनपेक्षा कमी असलेल्या व्हेरिएबल्सच्या व्हॅल्यूशी संबंधित फ्रिक्वेन्सीच्या तुलनेत जास्त फ्रिक्वेन्सी असेल. व्हिज्युअल आलेखामध्ये, अशा वितरणाची उजवी शेपटी लांब असते आणि शेपटी x अक्षाच्या सकारात्मक टोकाकडे (उच्च बाजू) निर्देशित करते म्हणून त्याला सकारात्मक तिरकस वितरण म्हणतात. हे अगदी तार्किक आहे की अशा वितरणाचा अर्थ > मध्यक> मोड आहे. खूप कमी लोकांना जास्त पगार मिळत असल्यामुळे कर्मचाऱ्यांच्या पगारात अनेकदा सकारात्मक बदल होतो. x च्या खालच्या मूल्यांशी संबंधित उच्च फ्रिक्वेन्सी x च्या उच्च मूल्यांच्या संगत फ्रिक्वेन्सीद्वारे ऑफसेट केल्या जात नाहीत.

सकारात्मक तिरकस वितरणाचे उदाहरण आणि आलेख:

X(चल)	14	16	18	20	22	24	26
f(फ्रेकेंव्सी)	2	8	17	10	7	4	2



- **असममित (नकारात्मक तिरकसपणा):(Negative Skewness):** असममित वितरणाला नकारात्मक तिरकस म्हटले जाते जर व्हेरिएबल्सच्या व्हॅल्यूजच्या व्हेरिएबल्समध्ये सरासरीपेक्षा कमी असलेल्या व्हेरिएबल्सशी संबंधित फ्रिक्वेन्सीच्या तुलनेत जास्त फ्रिक्वेन्सी असेल. अशा वितरणाची डावी शेपूट उजव्या शेपटापेक्षा लांब असते (म्हणून नकारात्मक तिरकस) कारण शेपूट x अक्षाच्या ऋण टोकाकडे निर्देशित करते. अशा डिस्ट्रिब्युशनमध्ये मोड > मीडियन> मीन व्यक्तींचे आयुष्य अनेकदा नकारात्मकरित्या तिरपे असते कारण x च्या उच्च मूल्यांशी संबंधित फ्रिक्वेन्सी x च्या कमी मूल्यांशी संबंधित फ्रिक्वेन्सीच्या तुलनेत जास्त असते.



- **तिरकसपणाचे परिपूर्ण उपाय (Absolute Measures of Skewness):**

खाली तिरपेपणाचे परिपूर्ण उपाय आहेत:

1. Skewness (S_k) = mean– median
2. Skewness (S_k) = mean – mode
3. Skewness (S_k) = (Q3 - Q2) - (Q2 - Q1)

मालिकेशी तुलना करण्यासाठी, आम्ही या निरपेक्ष मोजमापची गणना करत नाही सापेक्ष मापांची गणना करा ज्याला स्क्युनेसचे गुणांक(coefficient of skewness) म्हणतात. स्क्युनेसचे गुणांक हे एककांपासून स्वतंत्र असलेल्या शुद्धमोजमाप संख्या आहेत

- **तिरकसपणाचे सापेक्ष उपाय (Relative measure of Skewness):** दोन किंवा अधिकच्या विकृती दरम्यान वैध तुलना करण्यासाठी वितरण आम्हाला भिन्नतेचा वितरण प्रभाव दूर करावा लागेल. अशानिरपेक्ष विकृती मानकानुसार विभाजित करून निर्मूलन केले जाऊ शकते.विचलन सापेक्ष मोजण्यासाठी खालील महत्त्वाच्या पद्धती आहेत.

1. बीटा β आणि गॅमा γ स्क्युनेसचे गुणांक (β & γ Coefficient of Skewness)

Karl Pearson’s आणि skewness च्या गुणांकांची व्याख्या

दुसऱ्या आणि तिसऱ्या मध्यवर्ती क्षणांवर आधारित खाली केली आहे:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

हे तिरपेपणाचे मोजमाप म्हणून वापरले जाते. β_1 सममितीय वितरणासाठी शून्य असेल. β_1 तिरकसपणाचे उपाय असल्यामुळे तिरपेपणाची दिशा, म्हणजे सकारात्मक किंवा नकारात्मक बद्दल सांगत नाही. कारण μ_3 सरासरी पासून विचलनाच्या घनांची बेरीज असल्याने सकारात्मक किंवा ऋण असू शकते पण ${\mu^2}_3$ नेहमी सकारात्मक असतो. तसेच, μ_2 हा फरक नेहमीच सकारात्मक असतो. म्हणून,β_1 नेहमी सकारात्मक असेल. हा दोष दूर केला जातो जर आपण

कार्ल पियर्सनच्या गामा γ_1 गुणांकाची गणना करतो जे β_1 चे वर्गमूळ आहे

$$\gamma_1 = \pm\sqrt{\beta 1}$$

मग विकृतीचे चिन्ह μ_3 च्या मूल्यावर अवलंबून असेल की ते सकारात्मक किंवा नकारात्मक आहे. γ_1 चे मोजमाप म्हणून वापरणे उचित आहे.

तिरकसपणा

कार्ल पियर्सनचा स्क्युनेस गुणांक (स्क्युनेसचे पिअर्सोनियन गुणांक)

मध्य आणि मोडमधील चिन्हासह अंतर (+ किंवा - द्वारे दर्शविलेले) हे तिरपेपणाचे सूचक मानले जाऊ शकते. तथापि अशा मापाने म्हणजे मीन – मोड, भिन्न निसर्गाच्या वितरणांची तुलना करणे शक्य नाही.

उदाहरणार्थ: आपण नवीन जन्मलेल्या मानवी बाळांचे वजन आणि नवीन जन्मलेल्या सिंहाच्या पिल्लांच्या वजनाची तुलना करू शकत नाही. या दोन भिन्न गोष्टी आहेत आणि त्यामुळे त्यांची तुलना होऊ शकत नाही.

अशा परिस्थितीत तुलना करणे शक्य करण्यासाठी, आम्ही सरासरी आणि मोडमधील फरक मानक विचलनाने विभाजित करतो. कार्ल पियर्सनने ही पद्धत सुचवली आणि म्हणून त्याला पिअर्सोनियन गुणांक skewness म्हणून ओळखले जाते.

$Sk_p = \frac{Mean - Mode}{S.D.}$

Mode = (3Median-2 Mean) येथे, $-3 \le S_k \le 3$ व्यवहारात ते क्वचितच मिळते.

पिअर्सोनियन गुणांकाची वैशिष्ट्ये: Sk_p डेटा मूल्यांच्या युनिट्सपासून मुक्त आहे आणि मूल्यांच्या आकारापासून देखील मुक्त आहे. म्हणून, भिन्न आकाराच्या विविध वितरणांमध्ये तुलना करणे शक्य आहे.

- सममितीय वितरणासाठी, $Sk_p = 0$
- सकारात्मक तिरकस वितरणासाठी $Sk_p>0$
- नकारात्मक तिरकस वितरणासाठी $Sk_p < 0$

बहुतेक Sk_p मूल्ये -1 आणि 1 दरम्यान असतात; सर्व Sk_p मूल्य -3 आणि 3 च्या दरम्यान आहेत.

मोड अनिश्चित असल्यास आम्ही खालील वापरू शकतो:

मीन – मोड = 3(मीन- मीडियन);

Sk_pम्हणून परिभाषित केले जाऊ शकते $SK_p = 3\ (\frac{Mean - Mode}{S.D.})$

सोडवलेली उदाहरणे (Solved Problems)

उदाहरण-1.29: वितरणासाठी खालील डेटा आहे कंपन्यांचा नफा (लाख रुपयांत) Sk_p. शोधा.

नफा	10-20	20-30	30-40	40-50	50-60
फर्मचे क्र	12	18	25	10	7

(दिलेले = 11.7554)

उत्तर: 30-40 ने दर्शविलेला वर्ग हा मोडल क्लास आहे कारण त्याची कमाल फ्रेकेंव्सी आहे

f_m ही मोडल वर्ग = 25 ची फ्रेकेंव्सी आहे

f_1 ही प्री-मॉडल वर्गाची फ्रेकेंव्सी आहे = 18

f_2 ही पोस्ट मोडल वर्गाची फ्रेकेंव्सी आहे = 10

h = वर्ग-रुंदी = 10

L = मोडल वर्ग निम्न मर्यादा = 30

Mode $= L + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times h$

$= 30 + \frac{25-18}{2(25)-18-10} \times 10 \quad = 30 + \frac{70}{22} \ = 33.18$

वर्ग	फ्री-प्रमाण	फ्रेकेंव्सी	ui=(xi-35)/10	fiui	Fiui^2
10-20	12	15	-2	-24	48
20-30	18	25	-1	-18	18
30-40	25	35	0	0	0
40-50	10	45	1	10	10
50-60	7	55	2	14	28
	M=72			-18	104

$\bar{u} = \frac{\sum f_i u_i}{N} = \frac{-18}{72} = -0.25$

$\bar{x} = 35 + 10\bar{u} = 35 - 2.5 = 32.5$

$\sigma^2 = \frac{\sum f_i {u_i}^2}{N} - (\bar{u})^2 = \frac{104}{72} - (-0.25)^2$

$= 1.4444 - 0.0625 = \ 1.3819$

${\sigma_x}^2 = (10^2){\sigma_u}^2 = (100)(1.3819) = 138.19$

$\sigma_x = \sqrt{138.19} = 11.7554$

$Sk_p = \frac{Mean - Mode}{S.D.}$

$= \frac{32.5 - 33.18}{11.7554}$

$= \frac{-0.68}{11.7554}$

$= -0.0578\ ;\ Sk_p < 0$

नकारात्मक विकृती सूचित करते

बाउलीचे स्क्युनेसचे गुणांक: प्रो. बाउली यांनी विकृतीचे हे सापेक्ष उपाय विकसित केले. ज्यासाठी सूत्र दिले आहे:

$$Sk_b = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1)}$$

$$Sk_b = \frac{(Q_3 - 2Q_2 + Q_1)}{(Q_3 - Q_1)}$$

याचा आणखी उलगडा करूया; सममितीय वितरणासाठी, Sk_b हे 0 असे मानेल ते Q_1, Q_2 आणि Q_3 समान अंतरावर आहेत (कारण Q_3-Q_2=Q_2-Q_1 या प्रकरणात). सकारात्मक तिरकस वितरणासाठी Q_1, Q_3 च्या तुलनेत Q_2 च्या जवळ आहे कारण मूल्ये वितरणाच्या खालच्या भागाकडे (डावीकडे) जमा होतात. म्हणून ते त्याचे अनुसरण करते (Q_3-Q_2)>(Q_2-Q_1). हे सूचित करते की अंश सकारात्मक आहे आणि भाजक मूळतः सकारात्मक आहे, SK_b सकारात्मक आहे. याउलट, ऋणात्मक तिरकस वितरणासाठी, Q3, Q1 पेक्षा Q_2 च्या जवळ आहे, म्हणून Q_3-Q_2< Q_2-Q_1, म्हणून Skb चे मूल्य ऋण आहे (कारण अंश ऋणात्मक मूल्य गृहीत धरतो). Q1, Q2 आणि Q3 ची सापेक्ष स्थिती सकारात्मक आणि नकारात्मक तिरकस वितरणासाठी खालील आकृतीमध्ये दर्शविल्याप्रमाणे आहे:

सकारात्मक तिरकस (Positive skewness) (Q3-Q2) > (Q2-Q1)

नकारात्मक तिरकस (Negative skewness) ((Q3-Q2)< (Q2-Q1)

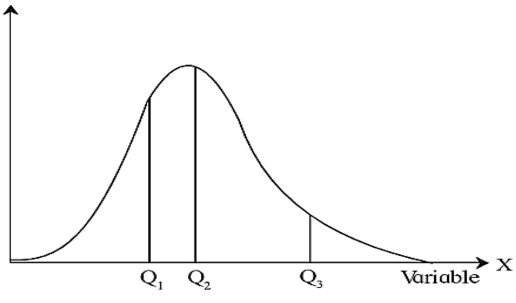


Fig 3.6 Positively Skewed Distribution

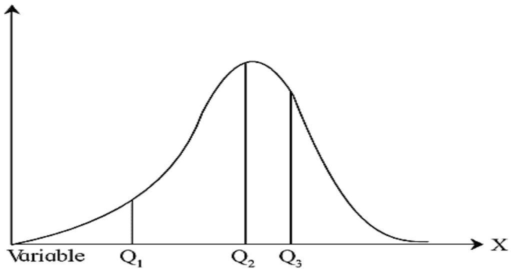


Fig 3.7 Negatively Skewed Distribution

टिप्पणी (Remarks):

1. Sk_b >0 हे सकारात्मक तिरकस वितरणाचे सूचक(Positively skewed distribution) आहे. त्याचप्रमाणे, Skb < 0 हे नकारात्मक तिरकस वितरणाचे सूचक (Negatively skewed distribution) आहे. Skb =0 सूचित सममितीय (symmetric) वितरण.
2. Skb -1 आणि 1 च्या दरम्यान आहे.
3. Skb ची गणना अगदी टोकावरील खुल्या वर्गांसह वितरणासाठी देखील केली जाऊ शकते. कारण हे खुले वर्ग Skb आधारित असलेल्या चतुर्थांशांवर परिणाम करत नाहीत. हे स्पष्ट करण्यासाठी पुढील उदाहरण पाहू.

सोडवलेली उदाहरणे (solved Problems)

उदाहरण-1.30: जर Q1 = 80, Q2 = 100, Q3 = 120, बाउलीचे तिरपेपणाचे गुणांक शोधा.

उत्तर:

$Sk_b = \frac{Q_3+Q_1-2Q_2}{Q_3-Q_1}$

$Sk_b = \frac{120+80-2(100)}{120-80}$

$Sk_b = \frac{200-(200)}{40} = \frac{0}{40} = 0$

Bowley's coefficient skewness गुणांक 0 आहे, म्हणून डेटा सममित आहे.

उदाहरण-1.31: कामगारांच्या साप्ताहिक वेतनाच्या खालील वितरणासाठी बाउलीचे स्क्युनेसच्या गुणांकाची गणना करा.

मजुरी खाली	300 च्या खाली	300-400	400-500	500-600	600-700	700 च्या वर
कामगार ची संख्या	5	8	18	35	27	7

उत्तर:

वर्ग	फ्रेकेंव्सी	प्रकारापेक्षा कमी वर्ग फ्रेकेंव्सी संचयी फ्रेकेंव्सी (less than cumulative frequency)
300 च्या खाली	5	5
300-400	8	13
400-500	18	31
500-600	35	66
600-700	27	93
700 च्या वर	7	100
एकूण	N=100	

Q1 = चे मूल्य $\frac{N}{4}$ th निरीक्षण

= २५ व्या निरीक्षणाचे मूल्य.

25 वे वर्गाचे निरीक्षण 400-500 आहे-

त्याची फ्रेकेंव्सी (f) 18 च्या बरोबरीची आहे आणि मागील वर्गाची संचयी फ्रेकेंव्सी 13 च्या बरोबरीची आहे.

$Q_1 = L + \frac{h\,(\frac{N}{4}-c.f.)}{f}$

$Q_1 = 400 + \frac{100\,(\frac{100}{4}-13)}{18}$

$Q_1 = 466.67$
Q2 = चे मूल्य $\frac{N}{2}$ th निरीक्षण
= 50 निरीक्षणाचे मूल्य.
50 वे वर्गाचे निरीक्षण वर्ग 500 – 600
f=35 आणि C.F. = 31 आहे
$Q_2 = L + \frac{h\,(\frac{N}{2} - c.f.)}{f}$
$Q_2 = 500 + \frac{100\,(\frac{100}{2} - 31)}{35}$
$Q_2 = 554.28$
Q3 = चे मूल्य$\frac{3N}{4}$ th निरीक्षण
75 व्या वर्गाचे निरीक्षण वर्ग 600 – 700
f = 27, C.F.= 66 चे आहे.
$Q_3 = L + \frac{h\,(\frac{3N}{4} - c.f.)}{f}$
$Q_3 = 600 + \frac{100\,(\frac{3(100)}{4} - 66)}{27}$
$Q_3 = 600 + \frac{100\,(75 - 66)}{27}$
$Q_3 = 663.33$
Bowley ‘s coefficient of skewness दिलेला आहे
$Sk_b = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$
$Sk_b = \frac{633.33 + 466.67 - 2(554.28)}{633.33 - 466.67}$

$Sk_b = \frac{1100 - 1108.56}{166.66} = -0.05$

उदाहरण-1.32: तिरकस वितरणासाठी मीन = 100, मीडियन= 98.5 आणि SD = 9., वितरणाचा मोड आणि पिअरसोनियन गुणांक (Skp) शोधा.
मीन – मोड = 3 (मीन - मीडियन)
100-मोड = 3(100-98.5)
मोड = 100 - 4.5 = 95.5

$$Sk_p = \frac{(Mean - Mode)}{S.D.}$$
$$= \frac{100 - 95.5}{9}$$
$$= \frac{4.5}{9}$$
$$= 0.5$$

Skp>0, वितरण सकारात्मक तिरकस (Positvely skewed)आहे.

उदाहरण-1.33: मोड 7 ने सरासरीपेक्षा मोठा आहे आणि भिन्नता 100 आहे. स्क्युनेसच्या गुणांकाची गणना करा. डेटा सकारात्मक किंवा नकारात्मक आहे का?

उत्तर:
मोड – मीन = 7
मध्य – मोड = −7
भिन्नता = 100 ---- दिल्याप्रमाणे.
एस.डी. = 10

$$Sk_p = \frac{(Mean - Mode)}{S.D.}$$
$$= \frac{-7}{10}$$
$$= -0.7$$

Skp < 0, वितरण ऋणात्मक तिरकस (Negatively skewed) आहे.

उदाहरण-1.34: जर अंकगणिताचा मीन = 214, मोड = 218, भिन्नता = 196, कार्ल पियर्सनचे विकृतीचे गुणांक आणि वितरणाचे स्वरूप शोधा.
उत्तर:

$$Sk_p = \frac{(Mean - Mode)}{S.D.}$$
$$= \frac{214-218}{\sqrt{196}}$$
$$= \frac{-4}{14}$$
$$= 0.2857 < 0$$

∴ वितरण ऋणात्मक तिरकस (Negatively skewed) आहे.

उदाहरण-1.35: फ्रेकेंव्सी वितरणासाठी, कमी चतुर्थक lower quartile35 आणि मध्यक(median)40 आहे. वितरण सममित असल्यास, वरचा चतुर्थक (upper quartile) शोधा.

उत्तर:
कमी चतुर्थक (lower quartile) = Q1 = 35
मीडियन(median)= 40 = Q2
Skb= 0 ---------- वितरण सममित आहे.
Skb = $\frac{Q_3+Q_1-2Q_2}{Q_3-Q_1}$
0 = $\frac{Q_3+Q_1-2Q_2}{Q_3-Q_1}$
0 = $Q_3 + Q_1 - 2Q_2$
$Q_3 = 2Q_2 - Q_1$
= 2(40)-35
=80-35
=45
वरचा चतुर्थांश (upper quartile) (Q3) = 45.

उदाहरण-1.36: जर Q1= 15, Q2= 21, Q3= 29 नंतर Bowley च्या skewnes च्या गुणांकाची गणना करा

उत्तर:

Skb $=\frac{Q_3+Q_1-2Q_2}{Q_3-Q_1}$

$Sk_b = \frac{29+15-2(21)}{29-15}$

$Sk_b = \frac{2}{14}$

$Sk_b = 0.143$

= 0.143 > 0, म्हणून वितरण धनात्मक तिरकस (Positively skewed) आहे.

उदाहरण-1.37: वितरणासाठी कार्ल पियर्सनचे स्क्युनेस गुणांक 0.64 आहे,मानक विचलन 13 आहे आणि सरासरी आहे 59.2 मोड आणि मध्य शोधा.

उत्तर: आम्ही दिले आहे Sk = 0.64, σ = 13 आणि mean = 59.2

म्हणून सूत्रे वापरून

$Sk_p = \frac{Mean- Mode}{S.D.}$

$0.64 = \frac{59.2-Mode}{13}$

Mode = 59.20 – 8.32 = 50.88

Mode = 3 Median – 2 Mean

50.88 = 3 Median – 2 (59.2)

Median = $\frac{50.88 +118.4}{3} = \frac{169.28}{3}$ = 56.42

सराव उदाहरणे

1. वितरणासाठी, सरासरी = 100 मोड = 127 आणि SD = 60. स्क्युनेस Skp चे पीअरसन गुणांक शोधा
2. वितरणाचा मध्य आणि भिन्नता अनुक्रमे 60 आणि 100 आहे. Skp असल्यास वितरणाचा मोड आणि मीडियनशोधा $Sk_p= -0.3$
3. डेटा सेटसाठी, वरच्या आणि खालच्या चतुर्थकांची बेरीज 100 आहे, वरच्या आणि खालच्या चतुर्थकांमधील फरक 40 आहे आणि मीडियन30 आहे. स्क्यूनेसचे गुणांक शोधा
4. 55 च्या बरोबरीच्या वरच्या चतुर्थक आणि 42 च्या मीडियनअसलेल्या डेटा सेटसाठी. वितरण सममित असल्यास, खालच्या चतुर्थकचे मूल्य शोधा
5. सूत्राद्वारे स्क्युनेसचे गुणांक मिळवा आणि वितरणाच्या स्वरूपावर टिप्पणी करा.

इंच मध्ये उंची	महिलांची संख्या
60 पेक्षा कमी	10
60-64	20
64-68	40
68-72	10
72-76	2

6. खालील निरीक्षणांच्या संचासाठी Skp शोधा.
7. 17,17,21,14, 15, 20, 19, 16, 13, 17, 18
8. तेरा (13) प्लॉट्समधून किलोमध्ये गव्हाच्या उत्पादनाच्या खालील निरीक्षणासाठी Sk_b ची गणना करा:
9. 5.24.6,3.5,4.8,5.5,4.7,3.6,3.5,4.2,3.5,3.6

10. फ्रेकेंव्सी वितरणासाठी Q3 – Q2= 90 आणि Q2 – Q1=120 Sk_b शोधा.
11. कार्ल पियर्सनचे स्क्युनेसचे गुणांक 1.28 आहे, त्याची सरासरी 164 आणि मोड 100 आहे ,SD शोधा.

1.9 मीन आणि मोडच्या दृष्टीने स्क्युनेसचे प्रकार (Types of Skewness in terms of Mean and Mode)

1. सकारात्मक तिरकसपणा- पॉसिटीव्ह स्केवणेस (Positive Skewness):
जर दिलेले वितरण डावीकडे हलवले आणि त्याची शेपटी उजवीकडे असेल, तर ते सकारात्मक तिरकस वितरण आहे. त्याला उजवे-तिरकस वितरण देखील म्हणतात. शेपटीला दुसऱ्या बाजूच्या डेटा बिंदूंपेक्षा वेगळ्या पद्धतीने वक्र निमुळता होणे म्हणून संबोधले जाते.
नावाप्रमाणेच, एक सकारात्मक तिरकस वितरण शून्यापेक्षा जास्त तिरकस मूल्य गृहीत धरते. दिलेल्या वितरणाचा तिरकसपणा उजवीकडे असल्याने, सरासरी मूल्य मध्यापेक्षा मोठे आहे आणि उजवीकडे हलते आणि मोड वितरणाच्या सर्वोच्च वारंवारतेवर येतो.

2. नकारात्मक तिरकसपणा- नेगेटिव्ह स्केवणेस (Negative Skewness):
दिलेले वितरण उजवीकडे आणि त्याची शेपटी डाव्या बाजूला हलवल्यास, ते ऋणात्मक तिरकस वितरण आहे. त्याला डावे-तिरकस वितरण देखील म्हणतात. ऋण तिरकस दर्शविणाऱ्या कोणत्याही वितरणाचे तिरकस मूल्य नेहमी शून्यापेक्षा कमी असते. दिलेल्या वितरणाचा तिरकसपणा डावीकडे आहे; म्हणून, सरासरी मूल्य मध्यापेक्षा कमी आहे आणि डावीकडे हलते आणि मोड वितरणाच्या सर्वोच्च वारंवारतेवर होतो.

3. स्क्युनेस मोजणे (Measuring Skewness): अनेक पद्धती वापरून स्क्युनेस मोजता येतो; तथापि, पिअर्सन मोड स्क्युनेस आणि पीअरसन मेडियन स्क्युनेस या दोन वारंवार वापरल्या जाणाऱ्या पद्धती आहेत. जेव्हा नमुना डेटाद्वारे मजबूत मोड प्रदर्शित केला जातो तेव्हा पिअरसन मोड स्क्युनेस वापरला जातो. डेटामध्ये एकाधिक मोड किंवा कमकुवत मोड समाविष्ट असल्यास, Pearson's Median skewness वापरले जाते.
पिअर्सन मोड स्क्युनेसचे सूत्र:

$$\text{स्क्युनेस (Skewness)} = \frac{\bar{x}-M_0}{s}$$

येथे

- $\bar{x}$= मध्य
- M_d = मोड
- s = मानक विचलन

पिअरसन मोड स्क्युनेसचे सूत्र:

$$\text{Skewness}=\frac{\overline{3x}-M_d}{s}$$

1.10 कुर्तोसिसची संकल्पना (Conept Of Kurtosis): आपल्याला केंद्रीय प्रवृत्ती, फैलाव आणि skewness माहित असल्यास, तरीही आपण वितरणाची संपूर्ण कल्पना मिळवू शकत नाही. यामध्ये उपायांव्यतिरिक्त, आम्हाला प्राप्त करण्यासाठी आणखी एक उपाय माहित असणे आवश्यक आहे. कुर्तोसिसची संकलपनेची मदत घेऊन वितरणाच्या आकाराबद्दल संपूर्ण कल्पना ज्याचा अभ्यास केला जाऊ शकतो. कर्टोसिस वितरणाच्या सपाटपणाचे मोजमाप देते. वितरणाच्या कर्टोसिसची डिग्री सामान्यच्या तुलनेत मोजली जाते. वक्र सामान्य वक्र पेक्षा जास्त शिखर असलेल्या वक्रांना "लेप्टोकर्टिक" "Leptokurtic".म्हणतात.
जे वक्र सामान्य वक्रांपेक्षा अधिक सपाट असतात"प्लॅटिकुर्टिक""Platykurtic" म्हणतात.
सामान्य वक्रला "मेसोकर्टिक" "Mesokurtic" म्हणतात.

कर्टोसिसचे उपाय(Measure of Kurtosis)

1. कार्ल पियर्सनचे कर्टोसिसचे उपाय (Karl Pearson's Measures of Kurtosis)

कर्टोसिसची गणना करण्यासाठी, व्हेरिएबलचे दुसरे आणि चौथे मध्यवर्ती क्षण वापरले जातात. यासाठी कार्ल पियर्सनने दिलेले खालील सूत्र वापरले आहे.

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

$$Y_2 = \beta_2 - 3$$

येथे,

- μ_2 = Second order central moment of distribution
- μ_4 = Fourth order central moment of distribution

1. If $\beta_2 = 3$ or $Y_2 = 0$, then curve is said to be mesokurtic
2. If $\beta_2 < 3$ or $Y_2 < 0$, then curve is said to be platykurtic
3. If $\beta_2 > 3$ or $Y_2 > 0$, then curve is said to be leptokurtic

उदाहरण 1.38: वितरणाच्या सरासरीबद्दल पहिले चार क्षण 0, 2.5, 0.7 आहेत.आणि 18.75. स्क्युनेस आणि कर्टोसिसचे गुणांक शोधा.

उत्तर: आमच्याकडे $\mu_1 = 0$, $\mu_2 = 2.5$, $\mu_3 = 0.7$ आणि $\mu_4 = 18.75$ आहेत.

परिमाणवाचक डेटाचे विश्लेषण

Skewness $\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(0.7)^2}{(2.5)^3} = 0.031$

Kurtosis $\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{18.75}{(2.5)^2} = 3$

$\beta_2 = 3$ curve is said to be mesokurtic

संदर्भ (References):

1. https://www.coursera.org/learn/stanford-statistics
2. https://www.mathworks.com/
3. https://brilliant.org/
4. https://onlinecourses.nptel.ac.in/noc24_ma30/preview

युनिट - 2
स्टॅटिस्टिकल मेथोड्स
(Statistical Methods)

अभ्यासक्रम निष्पत्ती (Course Outcome):

CO2: आर-प्रोग्रामिंग वापरून सांख्यिकी पद्धती लागू करा

घटक निष्पत्ती (Theory Learning Outcome):

1. किमान चौरस पद्धत वापरून सरळ रेषा आणि द्वितीय अंश बहुपदी फिट करा
2. कार्ल-पीअर्सन्स आणि स्पीयरमॅनच्या रँक पद्धती वापरून सहसंबंधाच्या सह-कार्यक्षमतेची गणना करा
3. दिलेल्या डेटासाठी रिग्रेशन रेषेचे समीकरण मिळवा.

2.1 सर्वात कमी चौरसांची पद्धत- मेथोड ऑफ लीस्ट Square (Method of Least Square): कमीत कमी चौरस पद्धत म्हणजे वक्रातून बिंदूंच्या ऑफसेट (अवशिष्ट भाग) च्या वर्गांची बेरीज कमी करून डेटा बिंदूंच्या संचासाठी सर्वोत्कृष्ट-फिटिंग वक्र किंवा सर्वोत्तम फिटची रेषा शोधण्याची प्रक्रिया आहे. ॲड्रिन-मायरे लॅजेंड्रे आणि कार्ल फ्रेडरिक यांनी गॉस एकमेकांपासून स्वतंत्रपणे कमीत कमी चौरसांची पद्धत विकसित केली. चला किमान चौरसांची मध्यवर्ती तत्त्वामागील कल्पना समजून घेऊया.

समजा मूल्ये (x1 , y1), (x2, y2), ..., (xn, yn) डेटामध्ये n जोड्यांचा समावेश आहे आणि समजा की दिलेल्या डेटाशी उत्तम बसणारी ओळ आहे खालीलप्रमाणे लिहिले आहे.

$\hat{Y} = a + b X$ (येथे, $\hat{Y}$ आहे Y कॅप म्हणून वाचिचे.) या समीकरणाला अंदाज समीकरण म्हणतात

अंदाज समीकरण स्थिरांकांची मूल्ये a आणि b वापरत आहे, अंदाजित मूल्य Y चे $\hat{Y}i = a + bXi$ द्वारे दिलेले आहेत , जेथे xi आहे स्वतंत्र व्हेरिएबलचे मूल्य आणि Yi आहे Y चे संबंधित अंदाजित मूल्य आहेत.

लक्षात ठेवा स्वतंत्र व्हेरिएबलचे निरीक्षण केलेले मूल्य yi हे अंदाजित मूल्यापेक्षा वेगळे आहे.

निरीक्षण मूल्यवान (yi) आणि अंदाज मूल्ये Y पूर्णपणे जुळत नाही कारण निरीक्षणे सरळ रेषेवर पडत नाहीत. निरीक्षण मूल्ये आणि अंदाजित मूल्य मध्ये असलेल्या फरकास त्रुटी किंवा अवशेष म्हणतात.

दिलेले मूल्य $y_1 - (a + bx_1)$, $y_2 - (a + bx_2)$, ..., $y_n - (a + bx_n)$ हे संबंधित अंदाजित मूल्यांमधून Y चे निरीक्षण केलेल्या मूल्यांचे विचलन आहे म्हणून त्यांना त्रुटी किंवा अवशेष म्हणतात. आपण लिहितो

$e_i = y_i - Y_1 = y_i - (a + bx_i)$, i = 1, 2, ..., n.

भौमितिकदृष्ट्या, अवशिष्ट ei, जे yi − (a + bxi), अनुलंब सूचित करते. निरीक्षण मूल्य (yi) आणि अंदाज मूल्य (Yi) दरम्यानचे अंतर (जे सकारात्मक किंवा नकारात्मक असू शकते) किमान वर्गाच्या पद्धतीचे खालील तत्त्व प्रमाणे सांगितले जाऊ शकते. सर्वोत्कृष्ट व्याख्या अशी आहे की अवशेषांच्या वर्गांची बेरीज, म्हणजेच अंदाजित y-मूल्यांचे निरीक्षण केलेल्या y-मूल्यांमधून विचलनांच्या वर्गांची बेरीज आहे. दुसऱ्या शब्दांत, The line of best fit निर्धारित करण्यासाठी स्थिर a आणि b प्राप्त केली जाते जेणेकरून किमान आहे.

$$\sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} (y_i - \hat{y_i})^2 = \sum_{i=1}^{n} [y_i - (a + bx_i)]^2$$

हे वापरून प्राप्त केलेली सरळ रेषा तत्त्वाला सर्वात कमी रिग्रेशन रेषा म्हणतात. लाक्षणिकरित्या, आपण लिहितो

$$S^2 = \sum_{i=1}^{n} (y_i - \hat{y_i})^2$$

चौरस त्रुटींची बेरीज म्हणून हे देखील असू शकते म्हणून लिहिले

$$S^2 = \sum_{i=1}^{n} [y_i - (a + bx_i)]^2$$

S^2 किमान असण्यासाठी आपल्याला a आणि b स्थिरांक ठरवायचे आहेत. लक्षा त घ्या की S2 हे a आणि b दोन्हीचे एक सतत आणि भिन्न कार्य आहे.
आपण S^2 मध्ये a च्या संदर्भात (ब हे स्थिर आहे गृहीत धरा) आणि b च्या संदर्भात (अ स्थिर गृहीत धरून)फरक करा. त्यानंतर आपण S^2 शून्यावर कमी करण्यासाठी या दोन्ही व्युत्पन्नांची समानता करतो . परिणामी,आपल्याला खालील दोन रेषीय समीकरणे दोन अज्ञात a आणि b मध्ये मिळतात .

$$\sum_{1=1}^{n} y_i = na + b\sum_{i=1}^{n} x_i$$

$$\sum_{1=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} {x_i}^2$$

जेव्हा आपण S^2 द्वारे ही दोन रेखीय समीकरणे सोडवतो, तेंव्हा a आणि b ची मूल्ये खालील प्रमाणे दिली जातात.

$$a = \bar{y} - b\bar{x}\,, b = \frac{cov(X,Y)}{{\sigma_x}^2}$$

Where

- $cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}$
- ${\sigma_x}^2 = \frac{1}{n}\sum_{i=1}^{n} {x_i}^2 - \bar{x}^2$

प्राप्त केलेल्या a आणि b ची मूल्ये रिग्रेशन समीकरणात वर दर्शविल्याप्रमाणे
Y = a + bX घालावीत तर आपणांस पुढील समीकरण मिळते.

$$y - \bar{y} = b(X - \bar{x})$$

सर्वात कमी चौरसांचे तत्त्व (Principle of least square): ग्राफिकल पद्धतीचा दोष आहे की काढलेली सरळ रेषा अनन्य असू शकत नाही परंतु कमीत कमी चौरसांचे तत्त्व स्थिरांकांना मूल्यांचा एक अद्वितीय संच प्रदान करते आणि म्हणून दिलेल्या डेटासाठी सर्वोत्तम फिट असलेली वक्र सुचवते. कमीत कमी चौरसाची पद्धत ही दिलेल्या डेटा पॉइंट्सद्वारे अद्वितीय वक्र बसवण्याची सर्वात पद्धतशीर प्रक्रिया आहे.
आपण काही सर्वोत्तम फिटिंग वक्रांचा विचार करू:

1. एक सरळ रेषा (A straight line)
2. दुसरा अंश वक्र(A second degree curve)
3. घातांक वक्र (The exponential curve) $y = ae^{bx}$.
4. वक्र $y = a x^n$

उदाहरण 2.1: कमीत कमी वर्गांच्या पद्धतीनुसार खाली दिलेल्या डेटाची सरळ रेषा शोधा:

X	5	10	15	20	25
Y	16	19	23	26	30

उत्तर:
सरळ रेषा y=ax+b आहे असे मानू
सामान्य समीकरणे आहेत
$\sum x$+5b = $\sum y$

$a\sum x^2 + b\sum x = \sum x\, y$

$\sum x , \sum y \sum x$y, $\sum x^2$ मोजणे

	x	y	x^2	xy
	5	16	25	80
	10	19	100	190
	15	23	225	345
	20	26	400	520
	25	30	625	750
Total	**75**	**114**	**1375**	**1885**

सामान्य समीकरणे आहेत

75a+5b=114 .....(eq.l)

1375a+75b=1885 ..... (eq.2)

b काढून टाकण्यासाठी, (1) 15 ने गुणा

1125a+75b=1710 .....(eq.3)

eq.2) - (eq.3) देते, 250 a=175 or a=0.7, hence b= 12.3

म्हणून, सर्वोत्तम फिटिंग लाइन आहे y=0.7x+12.3

उदाहरण 2.2: खाली दिलेल्या डेटाला सरळ रेषा बसवा. तसेच y चे मूल्य x=2.5 वर काढा

X	0	1	2	3	4
Y	1	1.8	3.3	4.5	6.3

उत्तर:

सरळ रेषा y=ax+b आहे असे मानू

सामान्य समीकरणे आहेत

$\sum x$+5b = $\sum y$

$a\sum x^2 + b\sum x$ = $\sum x\, y$

$\sum x$, $\sum y$ $\sum x$y, $\sum x^2$ मोजणे

	x	y	x^2	xy
	0	1	0	0
	1	1.8	1	1.8
	2	3.3	4	6.6
	3	4.5	9	13.5
	4	6.3	16	25.5
Total	**10**	**16.9**	**30**	**47.1**

(eq.2) आणि (eq.3) मध्ये बदलून, आपल्याला मिळते,

10a+5b=16.9.....(eq.l)

30a+10b=47.1 .....(eq.2)

b काढून टाकण्यासाठी, (1) 2ने गुणा

सोडवुन eq.(2)-eq.(l), आपणास मिळते,a=l .33, b= 0.72

त्यामुळे समीकरण आहे y =1.33x+0.72

y (at x=2.5) =1.33 (2.5) +0.72 = 4.045

उदाहरण 2.3: योग्य परिवर्तन करून, संबंध y=a + bx एका रेखीय फॉर्ममध्ये रूपांतरित करा आणि डेटा फिट करण्यासाठी समीकरण शोधा.

X	-4	1	2	3
Y	4	6	10	8

उत्तर: सरळ रेषा y=ax+b आहे असे मानू
सामान्य समीकरणे आहेत

$\sum x$+5b = $\sum y$

$a\sum x^2 + b\sum x$ = $\sum x\, y$

$\sum x$, $\sum y \sum x$y, $\sum x^2$ मोजणे.

	x	y	x	x^2	xy
	-4	4	-16	256	-64
	1	6	6	36	36
	2	10	20	400	200
	3	8	24	576	192
Total		**28**	**34**	**1268**	**364**

सामान्य समीकरणे आहेत

34a+1268b=364

4a+34b=28

सोडवुन eq.(2)-eq.(l), आपणास मिळते, a=5.90605, b= 0.12870

त्यामुळे समीकरण आहे y =5.90605 + 0.12870X

i.e. y= 5.90605 + 0.12870Xy

हे समीकरण वापरून आपल्याला मिळते y-0.12870 x=5.90605

2.2 कमीत कमी चौरस पद्धतीची पद्धत वापरून सरळ रेषेच्या दुसऱ्या बहुपदी $y=a+bx+cx^2$ चे फिटिंग (fitting of straight line second polynomial $y=a+bx+cx^2$ using method of least squares):

काहीवेळा डेटाच्या संचामध्ये द्वितीय अंश बहुपदी E(y| x) = α + β x + γ x^2 बसवण्याची इच्छा असते जेणेकरून त्याचा स्थिर बिंदू दिलेल्या प्रदेशात येईल. a, b, c, रिग्रेशन गुणांक α, β, आणि γ चे किमान चौरस अंदाज मोजण्यासाठी पद्धती दर्शविल्या जातात. एक उदाहरण दिले आहे ज्यामध्ये हा सिद्धांत डेटाच्या संचामध्ये पॅराबॉलिक रिग्रेशन फिट करण्यासाठी लागू केला जातो जेणेकरून स्वतंत्र व्हेरिएबलच्या दिलेल्या संलग्न आणि मर्यादित प्रदेशात, रिग्रेशन कार्य काटेकोरपणे वाढते आणि वरच्या दिशेने अवतल होते.

$y=a+bx+cx^2$ हा बिंदूंच्या (xi,yi) i=1,2,3..n च्या समुच्चयासाठी सर्वोत्तम फिट होणारा द्वितीय पदवी बहुपदी असू द्या ते आम्हाला कॉन्स्टंट a, b आणि c ठरवावे लागेल अंशतः E च्या संदर्भात व्युत्पन्न केल्याने आणि त्यांचे शून्यावर समीकरण केल्याने आपल्याला तीन सामान्य समीकरण मिळते. नंतर a,b आणि c चे मूल्य सोडवून

$\sum_{i=1}^{n} di^2 = \sum_{i=1}^{n}\left(\mathrm{yi} - (\mathrm{axi} + \mathrm{b})\right)^2$

By the principle of least squares, E is minimum

$\frac{\partial E}{\partial a} = 0 \;\; \frac{\partial E}{\partial b} = 0$

i.e.$2\sum(yi - (axi + c))(-xi) = 0$ &$2\sum(yi - (axi + c))(-1) = 0$

$\sum_{i=1}^{n}(xiyi - axi^2 - bxi) = 0$ & $\sum_{i=1}^{n}(yi - axi - b) = 0$

i.e. a$\sum_{i=1}^{n} xi^2$ +b$\sum_{i=1}^{n} xi = \sum_{i=1}^{n} xiyi$ ……(eq1)

and a$\sum_{i=1}^{n} xi + nb = \sum_{i=1}^{n} yi$

$\sum y=an+b\sum x+c\sum x^2$

उदाहरण 2.4: Calculate Fitting second degree parabola - Curve fitting using Least square method

X	1	2	3	4	5	6	7
Y	-5	-2	5	16	31	50	73

उत्तर:

The equation is $y=a+bx+cx^2$ and the normal equations are

$\sum y=an+b\sum x+c\sum x^2$

$\sum xy=a\sum x+b\sum x^2+c\sum x^3$

$\sum x^2y=a\sum x^2+b\sum x^3+c\sum x^4$

खालील सारणी वापरून मूल्यांची गणना केली जाते

x	y	x^2	x^3	x^4	xy	x^2y
1	-5	1	1	1	-5	-5
2	-2	4	8	16	-4	-8
3	5	9	27	81	15	45
4	16	16	64	256	64	256
5	31	25	125	625	155	775
6	50	36	216	1296	300	1800
7	73	49	343	2401	511	3577
---	---	---	---	---	---	---
$\sum x=28$	$\sum y=168$	$\sum x^2=140$	$\sum x^3=784$	$\sum x^4=4676$	$\sum xy=1036$	$\sum x^2y=644$

ही मूल्ये समीकरणांमध्ये ठेवा

$7a+28b+140c=168$

$28a+140b+784c=1036$

$140a+784b+4676c=6440$

ही ३ समीकरणे सोडवा,

$7a+28b+140c=168$……………..(1)

$28a+140b+784c=1036$................(2)

$140a+784b+4676c=6440$............(3)

समीकरणे (1) आणि (2) निवडा आणि व्हेरिएबल a काढून टाका

7a+28b+140c=16×4

28a+112b+560c=672 .... (4)

28a+140b+784c=1036×1

28a+140b+784c=1036 .... (5)

eq 4 –eq 5 आम्हाला मिळते -28b-224c=364

समीकरणे (1) आणि (3) निवडा आणि व्हेरिएबल a काढून टाका.

7a+28b+140c=168 x 20

140a+560b+2800c=3360 ....(6)

140a+784b+4676c=6440 x 1

140a+784b+4676c=6440 ....(7)

eq 6 –eq 7 आम्हाला मिळते -224b-1876c= -3080

28b-224c=364x

224b-179c=-2912 ....(8)

792c=2912

224b-1876c= - 3080 ....(9)

eq (8) –eq (9) we get

84c=168

C=2

Now From (6)

eq (1) X 4

From (4)

-28b-224c=-364

⇒-28b-224(2)=-364

⇒-28b-448=-364

⇒-28b=-364+448=84

⇒b=84/-28=-3

From (1)

7a+28b+140c=168

⇒7a+28(-3)+140(2)=168

⇒7a+196=168

⇒7a=168-196=-28

⇒a=-28/7=-4

Solution.

a=-4,b=-3,c=2

आता ही मूल्ये समीकरणात ठेवा $y=a+bx+cx^2$,

$y=-4-3x+2x^2$

उदाहरण 2.5: Calculate Fitting second degree parabola - Curve fitting using Least square method

X	1996	1997	1998	1999
Y	40	50	62	58

उत्तर:

The equation is $y=a+bx+cx^2$ and the normal equations are

$\sum y=an+b\sum x+c\sum x^2$

$\sum xy=a\sum x+b\sum x^2+c\sum x^3$

$\sum x^2y=a\sum x^2+b\sum x^3+c\sum x^4$

x	y	x=X-1998	x^2	x^3	x^4	xy	x^2y
1996	40	-2	4	-8	16	-80	160
1997	50	-1	1	-1	1	-50	50
1998	62	0	0	0	0	0	0
1999	58	1	1	1	1	58	58
2000	60	2	4	8	16	120	240
---	---	---	---	---	---	---	---
9990	270	0	10	0	34	48	508

ही मूल्ये समीकरणांमध्ये ठेवा

$270=5a+0b+10c$

$48=0a+10b+0c$

$508=10a+0b+34c$

ही 3 समीकरणे सोडवा,

$5a+10c=270$

$10b=48$

$10a+34c=508$

आता दिलेल्या समीकरणांचे मॅट्रिक्स फॉर्ममध्ये रूपांतर करा

$$\begin{bmatrix} 5 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 34 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 270 \\ 18 \\ 508 \end{bmatrix}$$

Now $A = \begin{bmatrix} 5 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 34 \end{bmatrix}, X = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ and $B = \begin{bmatrix} 270 \\ 18 \\ 508 \end{bmatrix}$

∴AX=B

∴X=A-1B

$|A| = 700$

Here,|A|=700≠0

∴A-1is possible.

$$Adj(A) = \text{Adj}\begin{bmatrix} 5 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 34 \end{bmatrix}$$

$$= \begin{bmatrix} 340 & 0 & -100 \\ 0 & 70 & 0 \\ -100 & 0 & 50 \end{bmatrix}$$

Now,A-1=$\frac{1}{|A|}$Adj(A)

Here,X=A-1×B

∴X=$\frac{1}{|A|}$Adj(A))×B

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \frac{410}{7} \\ \frac{24}{5} \\ \frac{-16}{7} \end{bmatrix}$$ **a** = $\frac{410}{7}$ **b** = $\frac{24}{5}$ **c** = $\frac{-16}{7}$

वरील समीकरणे सोडवून,

a = 58.5 **b** = 4.8 **c** = −2.28

$y = 58.5 + 4.8x - 2.28x^2$

सराव उदाहरणे

1. Fit a straight line y = a + bx using the following data

x	5	4	3	2	1
y	1	2	3	4	5

2. Fit a straight line y = a + bx using the following data

x	3	5	7	9	11
y	2.3	2.6	2.8	3.2	3.5

3. Fit a straight line to the following data on production.

Year	1996	1997	1998	1999	2000
Production	40	50	62	58	60

4. Fit a straight line to the following data on profit.

Year	1992	1993	1994	1995	1996	1997	1998	1999
Profit	38	40	65	72	69	60	87	95

5. Fit second degree parabola equation y=a+bx+cx2 using the following data.

x	12	10	8	6	4	2
y	6	5	4	3	2	1

2.3 कोव्यरिअन्स (Covariance): कोव्यरिअन्स हे दोन यादृच्छिक चलांमधील संबंधांचे मोजमाप आहे आणि ते एकत्र किती प्रमाणात बदलतात. किंवा आपण असे म्हणू शकतो की, दुसऱ्या शब्दांत, ते दोन व्हेरिएबलमधील बदल परिभाषित करते, जसे की एका व्हेरिएबलमधील बदल दुसऱ्या व्हेरिएबलमधील बदलाप्रमाणे असतो. जेव्हा

व्हेरिएबल्सचे रेखीय रूपांतर होते तेव्हा त्याचे स्वरूप राखण्याच्या फंक्शनचा हा गुणधर्म असतो. सहप्रसरण एककांमध्ये मोजले जाते, जे दोन चलांच्या एककांचा गुणाकार करून मोजले जाते

- **कोव्यरिअन्सचे प्रकार (Types of Covariance):**

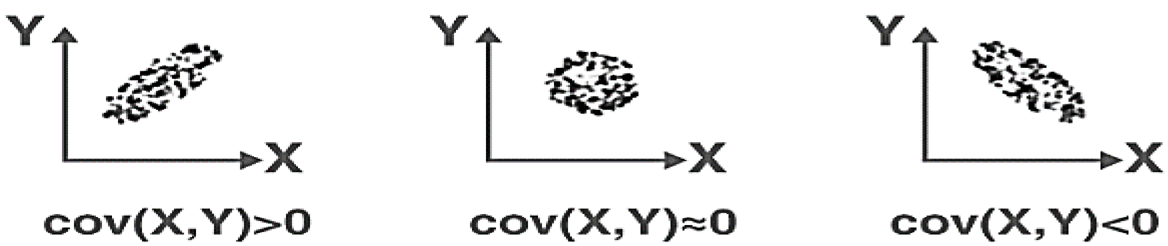
कोव्यरिअन्समध्ये सकारात्मक आणि नकारात्मक दोन्ही मूल्ये असू शकतात. यावर आधारित, त्याचे दोन प्रकार आहेत:

1. Positive covariance सकारात्मक कोव्यरिअन्स
2. Negative covariance नकारात्मक कोव्यरिअन्स

- **सकारात्मक कोव्यरिअन्स (Positive covariance):** कोणत्याही दोन व्हेरिएबल्ससाठी सहप्रसरण सकारात्मक असल्यास, याचा अर्थ, दोन्ही चल एकाच दिशेने जातात. येथे, व्हेरिएबल्स समान वर्तन दर्शवतात. याचा अर्थ, जर एका व्हेरिएबलची मूल्ये (मोठे किंवा कमी) दुसऱ्या व्हेरिएबलच्या मूल्यांशी जुळत असतील, तर ती सकारात्मक सहप्रसरणात आहेत असे म्हटले जाते.
- **नकारात्मक कोव्यरिअन्स (Negative covariance):** कोणत्याही दोन चलांसाठी सहप्रसरण ऋण असेल तर, याचा अर्थ, दोन्ही चल विरुद्ध दिशेने जातात. हे सकारात्मक सहप्रवर्तनाचे विरुद्ध प्रकरण आहे, जेथे एका व्हेरिएबलची मोठी मूल्ये दुसऱ्या व्हेरिएबलच्या कमी मूल्यांशी संबंधित असतात आणि त्याउलट.
- **कोव्यरिअन्स सूत्र (Covariance Formula):** Covariance सूत्र हे सांख्यिकीय सूत्र आहे, दोन चलांमधील संबंधांचे मूल्यांकन करण्यासाठी वापरले जाते. दोन चलांमधील फरकांमधील संबंध जाणून घेण्यासाठी हे एक सांख्यिकीय मोजमाप आहे. X आणि Y हे कोणतेही दोन व्हेरिएबल्स आहेत, ज्यांचा संबंध मोजावा लागेल असे समजू. अशा प्रकारे या दोन व्हेरिएबल्सचे सहप्रसरण Cov(X,Y) द्वारे दर्शविले जाते. लोकसंख्येतील सहप्रवाह आणि नमुना सहप्रवाह या दोन्हीसाठी सूत्र खाली दिले आहे.
- Cov (x,y) = $\frac{\sum(xi-x)(yi-y)}{N}$

Where,

- xi = data value of x
- yi = data value of y
- $\bar{x}$ = mean of x
- $\bar{y}$ = mean of y
- N = number of data values.



Covariance of X and Y

1. जर cov(X, Y) शून्यापेक्षा जास्त असेल, तर आपण असे म्हणू शकतो की कोणत्याही दोन व्हेरिएबल्ससाठी सहप्रसरण धनात्मक आहे आणि दोन्ही चल एकाच दिशेने जातात.
2. जर cov(X, Y) शून्यापेक्षा कमी असेल, तर आपण असे म्हणू शकतो की कोणत्याही दोन व्हेरिएबल्ससाठी सहप्रसरण ऋण आहे आणि दोन्ही चल विरुद्ध दिशेने जातात.
3. जर cov(X, Y) शून्य असेल, तर आपण असे म्हणू शकतो की दोन चलांमध्ये कोणताही संबंध नाही.

कोव्हरिअन्सचे उदाहरण - (Covariance Example):

Below example helps in better understanding of the covariance of among two variables**.**

उदाहरण 2.6: Calculate the coefficient of covariance for the following data:

X	2	8	18	20	28	30
y	5	12	18	23	45	50

उत्तर:

निरीक्षणांची संख्या = 6

X चा मीन = 17.67

Y चा मीन = 25.5

Cov(X, Y) = (⅙) [(2 – 17.67)(5 – 25.5) + (8 – 17.67)(12 – 25.5) + (18 – 17.67)(18 – 25.5) + (20 – 17.67)(23 – 25.5) + (28 – 17.67)(45 – 25.5) + (30 – 17.67)(50 – 25.5)]

= 157.83

2.4 Concept of correlation, Types of Correlation

Concept of correlation: दोन चलांमधील काही संबंध किंवा संघटना आहे का ते शोधणे. "एक सहसंबंध म्हणजे सहवास किंवा नातेसंबंधाचे मोजमाप". जर आपण निरीक्षण केले Bivariate डेटामध्ये एका व्हेरिएबलमधील बदल इतर व्हेरिएबलमधील बदलांसह आहेत मग दोन व्हेरिएबल्स परस्परसंबंधित असल्याचे म्हटले जाते.या प्रकरणात आम्ही म्हणतो की दोन अंतर्निहित चलांमधील परस्परसंबंध आहे

उदाहरणार्थ,

1. बुद्धिमत्ता भाग (IQ) आणि विद्यार्थी चे गुण
2. वस्तूची मागणी आणि किंमत

Covarince: सहविभाजन हे संयुक्त भिन्नतेचे एक माप आहे दोन व्हेरिएबल्स दरम्यान. जर (x1, y1), (x2, y2),) x आणि y च्या मूल्यांच्या n क्रमबद्ध जोड्या आहेत,नंतर X आणि Y मधील सहप्रसरण द्वारे परिभाषित केले जाते

2.4 कोरिलेशन (Correlation): कोरिलेशन जोडणीचा संदर्भ देते, कोरिलेशन विश्लेषणामध्ये आम्ही कनेक्शन किंवा संबंधाचा अभ्यास करतो दोन किंवा अधिक व्हेरिएबल्स दरम्यान. जर दोन व्हेरिएबल अशा प्रकारे बदलत असतील तर एका व्हेरिएबलमधील बदल इतर व्हेरिएबलमधील बदलांसह, हे चल परस्परसंबंधित असल्याचे म्हटले जाते.

उदाहरणार्थ,वर्गातील विद्यार्थ्यांची उंची आणि वजन यांच्यात काही संबंध असतो.कुटुंबाची कमाई आणि आलिशान वस्तूंवर खर्च केलेली रक्कम, इन्सुलिनचा डोस आणि डोस यांच्यातील संबंध रक्तातील साखर, इ. व्हेरिएबल्सची संख्या 'n' असू शकते जी एकमेकांवर आणि नातेसंबंधांवर परिणाम करू शकतात त्या सर्वांमध्ये अस्तित्वात असू शकते उदाहरणार्थ X, Y आणि Z हे तीन व्हेरिएबल्स आहेत, X Z शी संबंधित असू शकता आणि Y आणि Z दोन्ही Y शी संबंधित असू शकतात, म्हणून आम्ही सांगू शकतो की कोरिलेशन हे दोन किंवा अधिक चलांमधील संबंध शोधण्यासाठी एक सांख्यिकी साधन आहे.

- **कोरिलेशनची वैशिष्ट्ये (Characteristics Corrrlation)**

1. जरी कोरिलेशन विश्लेषण व्हेरिएबल्समधील संबंधांची डिग्री स्थापित करते परंतु ते अयशस्वी होते. कारण-प्रभाव संबंधांवर प्रकाश टाका.

2. व्हेरिएबल्समधील परस्परसंबंधाचे अस्तित्व संयोगामुळे असू शकते, विशेषतः जेव्हा नमुना घेतलेला असतो संख्येने लहान. जर आपण काही डेटा पॉइंट्स किंवा विशिष्ट व्हेरिएबलमधील बदलाचे काही निरीक्षण लिहितो निरीक्षणामध्ये काही नमुना किंवा संबंध असण्याची शक्यता आहे जी कदाचित हेतुपुरस्सर नसेल त्याचप्रमाणे भिन्न डेटा संच किंवा भिन्न व्हेरिएबल्सच्या निरीक्षणांमध्ये देखील हे शक्य आहे वास्तविक परिस्थितीत खरे नाही.

उदाहरणार्थ, आम्ही एनसीआरमध्ये विकल्या गेलेल्या बाईकच्या संख्येचा डेटा संच आणि पावसाचे प्रमाण गोळा करतो.

Year	Number of Bikes sold in NCR (In lakhs) NCR मध्ये विकल्या गेलेल्या बाईकची वर्षांची संख्या	Total Rainfall in NCR (MM) NCR मध्ये एकूण पाऊस (MM)
2001	25	120
2002	35	140
2003	45	160
2004	55	180
2005	65	200
2006	75	220
2007	85	240
2008	95	260

2.4.1 कोरिलेशनचे प्रकार (Types of Correlation): विद्वानांनी विविध प्रकारांमध्ये कोरिलेशन परिभाषित केले आहेत. मुळात ते सर्व तीन प्रकारात परिभाषित केले जाऊ शकतात-

प्रकार 1: ह्या प्रकारा मध्ये निरीक्षणातील बदलाची कोरिलेशन दिशा महत्त्वाची आहे, ती अधिक एकदा समजू शकतेआम्ही सकारात्मक आणि नकारात्मक कोरिलेशनमध्ये फरक करतो.

1.सकारात्मक कोरिलेशन (Positive Correlation): X आणि Y ही दोन व्हेरिएबल्स घेऊ या दोन्ही व्हेरिएबल्समधील बदलाची दिशा अशा प्रकारे असेल की जर X सरासरीने कमी झाले तर Y सुद्धा कमी होते आणि X वाढल्यास Y ची सरासरी देखील वाढते. या परस्परसंबंधाला सकारात्मक कोरिलेशन म्हणतात. उदाहरणार्थ:

उदाहरण 1

X	10	12	11	18	20
y	15	20	22	25	37

उदाहरण 2

X	80	70	60	40	30
Y	50	45	30	20	10

दोन्ही प्रकरणांमध्ये आपण पाहतो की X आणि Y मधील बदल एकाच दिशेने आहे. सकारात्मक कोरिलेशनची उदाहरणे उंची आणि वजन, पाण्याचा वापर आणि तापमान इत्यादी असू शकतात.

2. नकारात्मक कोरिलेशन (Negative Correlation): त्याचप्रमाणे नकारात्मक कोरिलेशनत सकारात्मक कोरिलेशन बदलाची दिशा महत्त्वाची असते परंतु विरुद्ध असते दिशा. जर आपण X आणि Y व्हेरिएबल्सचे समान उदाहरण घेतले आणि पाहिले की X वाढत आहे का आणि वर Y जर सरासरी कमी होत असेल किंवा X कमी होत असेल आणि Y सरासरीने वाढत असेल, तर यासह चल कोरिलेशन नकारात्मक कोरिलेशन म्हणतात,

उदाहरणार्थ-

परिस्थिती 1

X	20	30	40	60	80
y	40	30	22	15	16

परिस्थिती 2

X	100	90	60	40	30
y	10	20	30	40	50

वरील दोन्ही प्रकरणांमध्ये आपण पाहतो की X आणि Y मधील बदल विरुद्ध दिशेने आहे. मागणीचा कायदा आहे नकारात्मक कोरिलेशनचे उदाहरण.

प्रकार 2: ह्या प्रकारा मध्ये अभ्यास केलेल्या चलांची कोरिलेशन संख्या महत्वाची आहे. चल संख्या ठरवतात की ते साधे किंवा अनेक आहेत, हे समजून घेण्यासाठी आपण एकामागून एक चर्चा करूया-

1. साधा कोरिलेशन (Simple Correlation):

जेव्हा एखाद्या समस्येमध्ये आपण फक्त दोन चलांचा अभ्यास करतो, तेव्हा समस्येला एक साधी कोरिलेशन समस्या म्हटले जाते. च्या साठी उदाहरणार्थ, जर आपण एखाद्या क्षेत्रातील गव्हाचे उत्पादन आणि त्याच क्षेत्रातील पावसाचे प्रमाण अभ्यासले तर ते होईल साध्या सहसंबंधाचे उदाहरण मानले जाते.

2. एकाधिक कोरिलेशन (Multiple Correlation):

जेव्हा दोनपेक्षा जास्त चलांचा अभ्यास केला जातो तेव्हा समस्या अनेक कोरिलेशनची असते. उदाहरणार्थ जर आम्ही एखाद्या भागातील भाताचे उत्पादन, पावसाचे प्रमाण आणि वापरल्या जाणाऱ्या खतांचे प्रमाण यांचा अभ्यास करा. समान क्षेत्र, ते या सहसंबंधाची समस्या म्हणून मानले जाईल.

व्हेरिएबल्सच्या प्रभावशाली स्वरूपानुसार ते दोन प्रकारचे असते-

3. आंशिक कोरिलेशन (Partial Correlation): आंशिक कोरिलेशन येतो जेथे समस्येमध्ये आपण दोनपेक्षा जास्त चल ओळखतो परंतु केवळ विचारात घेतो.इतर व्हेरिएबल्स स्थिर मानून, एकमेकांवर प्रभाव टाकणारे दोन चल.उदा. गहू उत्पादनाच्या समस्येत, पर्जन्यमान आणि तापमान यांचे परस्परसंबंध विश्लेषण केले तर गहू उत्पादन आणि पर्जन्यमान यांच्या दरम्यान आणि ठराविक सरासरी दररोजच्या कालावधीपर्यंत मर्यादित आहे तापमान अस्तित्वात होते.

4. एकूण कोरिलेशन (Total Correlation): जेव्हा आपण सर्व विद्यमान व्हेरिएबल्सचा अभ्यास करतो तेव्हा ही संपूर्ण परस्परसंबंधाची समस्या असते. तथापि ते सामान्यपणे आहे अशक्य किंवा अशक्य जवळ.

प्रकार 3: कोरिलेशनमध्ये आपण बदलाच्या गुणोत्तरावर आधारित व्हेरिएबलमधील संबंध वेगळे करतो. चे प्रमाण दोन व्हेरिएबल्स ठरवतात की दोन्ही व्हेरिएबल्स रेखीय सहसंबंधित आहेत की नॉन-रेखीय.

1. रेखीय कोरिलेशन (Linear Correlation): रेखीय कोरिलेशन बदलाच्या गुणोत्तरावर आणि अभ्यासाधीन चलांमधील सातत्य यावर आधारित आहे. गुणोत्तर स्थिर असताना व्हेरिएबल्स रेखीय सहसंबंधित असल्याचे म्हटले जाते. दुसऱ्या शब्दांत जर ची पदवी एका व्हेरिएबलमधील बदल आणि इतर व्हेरिएबलमधील बदलाचे प्रमाण समान गुणोत्तरात परिणाम करतात, नंतर कोरिलेशन रेखीय असल्याचे म्हटले जाते.

जर आपण आलेख काढला तर रेखीय संबंध असलेल्या चलांची नेहमी सरळ रेषा तयार होईल.

उदाहरणार्थ: खाली रेखीय कोरिलेशनची उदाहरणे दिली आहेत-

X = 10	20	30	40	50	60	70	80
Y = 70	140	210	280	350	420	490	560

2. नॉन-रेखीय कोरिलेशन (Non-Linear Correlation): जेव्हा एका व्हेरिएबलमधील बदलाचे प्रमाण मधील बदलाच्या प्रमाणाशी स्थिर गुणोत्तर धरत नाही इतर चल, ते नॉनलाइनर कोरिलेशन असल्याचे म्हटले जाते. कोरिलेशन रेखीय नॉन-रेखीय उदा.- पाऊस दुप्पट झाल्याने तांदूळ गव्हाचे उत्पादन दुप्पट वाढणार नाही. या कोरिलेशन हा चलांमधील वक्र संबंध म्हणून देखील ओळखला जातो.

कोरिलेशनचा अभ्यास करण्याच्या विविध पद्धती (Various techniques of studying Correlation)

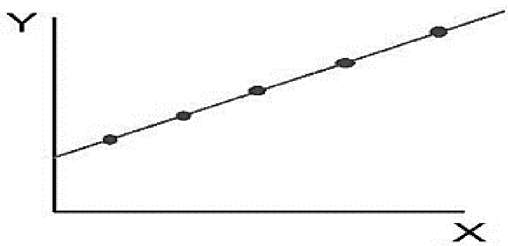
कोरिलेशनचा अभ्यास करण्यासाठी खालील तंत्रे वापरली जातात:

1. कोरिलेशनचे स्कॅटर डायग्राम तंत्र (Scatter diagram technique of correlation)
2. कार्ल पियर्सनचा कोरिलेशन गुणांक (Karl Pearson’s Coefficient of correlation)

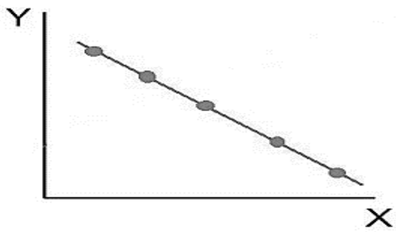
3. स्पिअरमॅनचा रँक कोरिलेशन (Spearman's Rank Correlation)

2.5 स्कॅटर किंवा डॉट डायग्राम पद्धतीची व्याख्या (Scatter or Dot Diagram Method Definition): स्कॅटर डायग्राम पद्धत ही दोन व्हेरिएबल्समधील परस्परसंबंधाचा अभ्यास करण्याची सर्वात सोपी पद्धत आहे ज्यामध्ये व्हेरिएबलच्या प्रत्येक जोडीची मूल्ये एका आलेखावर ठिपक्यांच्या स्वरूपात प्लॉट केली जातात आणि त्याद्वारे त्यांच्या संख्येइतके बिंदू प्राप्त होतात. निरीक्षणे मग अनेक बिंदूंचे विखुरलेले प्रमाण पाहून, परस्परसंबंधाची डिग्री निश्चित केली जाते. ज्या प्रमाणात व्हेरिएबल्स एकमेकांशी संबंधित आहेत ते बिंदू चार्टवर विखुरलेल्या पद्धतीवर अवलंबून असतात. प्लॉट केलेले पॉइंट्स चार्टवर जितके जास्त विखुरले जातील तितके व्हेरिएबल्समधील परस्परसंबंधाचे प्रमाण कमी असेल. प्लॉट केलेले बिंदू रेषेच्या जितके अधिक जवळ असतील, तितकी कोरिलेशनची डिग्री जास्त असेल. कोरिलेशनची डिग्री "r" द्वारे दर्शविली जाते. खालील प्रकारचे स्कॅटर डायग्राम व्हेरिएबल X आणि व्हेरिएबल Y यांच्यातील कोरिलेशनच्या डिग्रीबद्दल सांगतात.

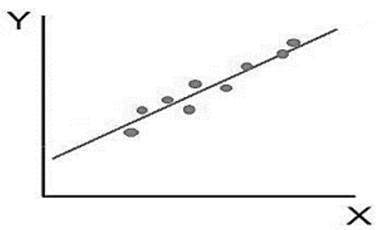
1. Perfect positive corelation परफेक्ट पॉझिटिव्ह सहसंबंध (r=+1): जेव्हा सर्व बिंदू खालच्या डाव्या कोपऱ्यापासून वरच्या उजव्या कोपऱ्यापर्यंत जाणाऱ्या सरळ रेषेवर असतात तेव्हा कोरिलेशन पूर्णपणे सकारात्मक असल्याचे म्हटले जाते.



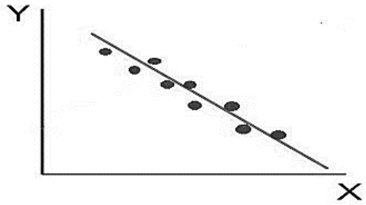
2. Perfect negative corelation परफेक्ट नकारात्मक कोरिलेशन (r=-1): जेव्हा सर्व बिंदू वरच्या डाव्या कोपऱ्यापासून खालच्या उजव्या कोपर्यात पडणाऱ्या सरळ रेषेवर असतात, तेव्हा व्हेरिएबल्सला ऋणात्मक कोरिलेशन असल्याचे म्हटले जाते.



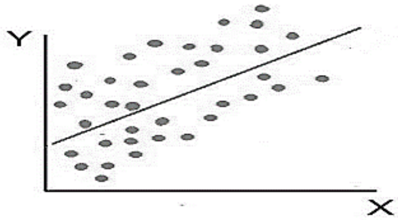
3. High Degree of +Ve Correlation (r= + High): +**Ve** कोरिलेशनची उच्च पदवी (**r= +** उच्च): जेव्हा प्लॉट केलेले बिंदू अरुंद पट्ट्याखाली येतात तेव्हा कोरिलेशनची डिग्री जास्त असते आणि जेव्हा ते खालच्या डाव्या कोपऱ्यापासून वरच्या बाजूस वाढणारी प्रवृत्ती दर्शवतात तेव्हा ते सकारात्मक असल्याचे म्हटले जाते. उजव्या हाताचा कोपरा.



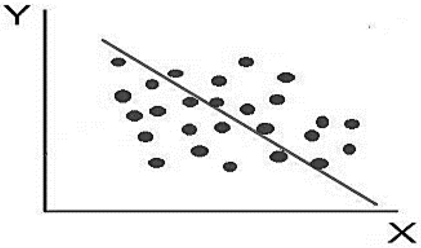
4.High Degree of –Ve Correlation (r= – High): –Ve कोरिलेशनची उच्च पदवी (r= – उच्च): जेव्हा प्लॉट केलेला बिंदू अरुंद बँडमध्ये पडतो आणि वरच्या डाव्या-कोपऱ्यापासून खालच्या उजव्या कोपर्यापर्यंत घसरणारी प्रवृत्ती दर्शवतो तेव्हा नकारात्मक सहसंबंधाची डिग्री जास्त असते.



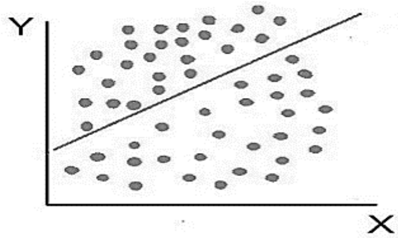
5. Low degree of +Ve Correlation (r= + Low): +Ve सहसंबंधाची कमी पदवी (r= + Low): जेव्हा बिंदू आलेखावर जास्त प्रमाणात विखुरलेले असतात आणि वरच्या उजव्या हाताच्या कोपऱ्यात डाव्या हाताच्या कोपऱ्यात वाढणारी प्रवृत्ती दर्शवतात तेव्हा व्हेरिएबल्समधील कोरिलेशन कमी असतो परंतु सकारात्मक असल्याचे म्हटले जाते.



6. Low Degree of –Ve Correlation (r= + Low): -Ve सहसंबंधाची निम्न पदवी (r= + Low): जेव्हा बिंदू आलेखावर विखुरलेले असतात आणि वरच्या डाव्या कोपऱ्यापासून खालच्या उजव्या कोपऱ्यापर्यंत घसरण्याची प्रवृत्ती दर्शवते तेव्हा कोरिलेशनची डिग्री कमी आणि ऋण असते.



7. No Correlation (r= 0): कोणताही कोरिलेशन नाही (r= 0): जेव्हा बिंदू आलेखावर अव्यवस्थितपणे विखुरलेले असतात आणि कोणताही विशिष्ट नमुना दर्शवत नाहीत तेव्हा व्हेरिएबल असंबंधित असल्याचे म्हटले जाते. येथे कोरिलेशन अनुपस्थित आहे आणि म्हणून r = 0.



अशाप्रकारे, दिलेल्या व्हेरिएबल व्हॅल्यूच्या प्रत्येक जोडीसाठी ठिपके प्लॉट करून व्हेरिएबल्समधील नातेसंबंधाचा अभ्यास करण्यासाठी स्कॅटर डायग्राम पद्धत ही सर्वात सोपी उपकरण आहे. ज्या तक्त्यावर ठिपके असतात त्याला डॉटोग्राम असेही म्हणतात

2.6 कोरिलेशन संकल्पना (Concept of Correlation):

आपण आधीच शिकलो आहोत की दोन चलांमधील संबंधांची ताकद आणि दिशा मोजण्यासाठी सहसंबंध वापरला जातो. सांख्यिकीमध्ये, सहसंबंध दोन परिमाणवाचक चलांमधील संबंध दर्शवतो. असे गृहीत धरले जाते की हा संबंध रेखीय आहे. म्हणजेच, एक व्हेरिएबल इतर व्हेरिएबलमध्ये वाढ किंवा घटण्याच्या प्रत्येक युनिटसाठी एका निश्चित रकमेने वाढतो किंवा कमी होतो. खालील प्रत्येक उदाहरणातील दोन व्हेरिएबल्समधील संबंध विचारात घ्या.

1. उत्पादनाची जाहिरात आणि विक्री. (सकारात्मक कोरिलेशन)
2. प्राथमिक शाळेतील विद्यार्थ्यांची उंची आणि वजन. (सकारात्मक कोरिलेशन)
3. खताची मात्रा आणि पीक उत्पादनाची रक्कम. (सकारात्मक कोरिलेशन)
4. व्यायाम आणि वजन कमी करण्याचा कालावधी. (सकारात्मक कोरिलेशन)
5. वस्तूची मागणी आणि किंमत. (सकारात्मक कोरिलेशन)
7. वस्तूचा पुरवठा आणि किंमत. (नकारात्मक कोरिलेशन)
8. गैरहजेरीचे दिवस (शाळेत) आणि परीक्षेतील कामगिरी. (नकारात्मक कोरिलेशन)
9. जितके जास्त जीवनसत्त्वे वापरतात, तितकी त्याची कमतरता असते. (नकारात्मक कोरिलेशन)

कोरिलेशन गुणांक दोन व्हेरिएबल्समधील संबंध मोजतात परंतु जेव्हा दुसऱ्या व्हेरिएबलचे मूल्य ओळखले जाते किंवा दिले जाते तेव्हा ते एका व्हेरिएबलचे मूल्य निर्धारित करू शकत नाही.

एका व्हेरिएबलच्या मूल्याचा अंदाज लावण्यासाठी इतर व्हेरिएबलच्या दिलेल्या मूल्याला रिग्रेशन म्हणतात. रीग्रेशन हे व्हेरिएबल्समधील संबंध तपासण्यासाठी सांख्यिकीय साधन आहे. मूल्याचा अंदाज लावण्यासाठी आणि परिणामास कारणीभूत घटक ओळखण्यासाठी हे वारंवार वापरले जाते. कार्ल पीअरसन कोरिलेशनची गुणांक परिभाषित करतात ज्याला Pearson's Product Moment सहसंबंध गुणांक म्हणून ओळखले जाते कार्ल फ्रेडरिक गॉस यांनी दोन किंवा अधिक चलांमधील संबंधांचे सर्वोत्तम वर्णन करणारी रेषीय समीकरण शोधण्यासाठी Least Squares Method म्हणून ओळखली जाणारी पद्धत विकसित केली. आर.ए. फिशरने गॉस आणि पियर्सन यांच्या कार्याला एकत्रित करून रेखीय प्रतिगमनातील किमान चौरस अनुमानाचा संपूर्ण सिद्धांत विकसित केला. फिशरच्या कार्यामुळे, रेखीय प्रतिगमनाचा वापर अंदाज आणि परस्परसंबंध समजून घेण्यासाठी केला जातो.

टीप: काही सांख्यिकीय पद्धती प्रयत्न करतात

अज्ञात प्रमाणाचे मूल्य निर्धारित करण्यासाठी, जे पॅरामीटर किंवा यादृच्छिक व्हेरिएबल असू शकते. या उद्देशासाठी वापरल्या जाणाऱ्या पद्धतीला जर अज्ञात परिमाण हे पॅरामीटर असेल तर अंदाज म्हणतात आणि अज्ञात परिमाण व्हेरिएबल असल्यास अंदाज म्हणतात.

कार्ल पियर्सनचा कोरिलेशन गुणांक (Karl Pearson's Coefficient of correlation):

जेव्हा आपण दोन व्हेरिएबल्सचा अभ्यास करतो तेव्हा काही तात्काळ प्रश्न मनात येतात, प्रथम काही संबंध आहे का जर व्हेरिएबल्समध्ये काही संबंध असेल तर या दोन व्हेरिएबल्सच्या वर्तनामध्ये दुसरा हे सांख्यिकीयदृष्ट्या पुरेसे महत्त्वाचे आहे की नाही हे तिसरे संबंध (+ve किंवा –ve) कोणत्या प्रकारचे आहे.व्हेरिएबल्स आणि चौथा म्हणजे एका व्हेरिएबलच्या मूल्याचा दुसऱ्या व्हेरिएबलमधील बदलाच्या आधारे अंदाज लावला जाऊ शकतो.चल? जर या प्रश्नांची उत्तरे अचूक दिली गेली तर आपण दोघांच्या वर्तनाचे मोजमाप आणि अंदाज लावू शकतो.चल सहसंबंध विश्लेषण पहिल्या तीन प्रश्नांची उत्तरे देते आणि शेवटच्या प्रश्नाचे उत्तर दिले जाते रिग्रेशन विश्लेषण. हे केवळ दोन चलांमधील नातेसंबंधाच्या सामर्थ्याबद्दलच नाही बदलाची दिशात्यांच्याबद्दल देखील सांगते . हे सराव मध्ये मोठ्या प्रमाणावर वापरले जाते आणि Pearson’s coefficient of Corelation म्हणून प्रसिद्ध आहे .ते दर्शविले जाते 'r' द्वारे. ही एक गणितीय पद्धत आहे जी चलांमधील रेखीय संबंधांची डिग्री मोजण्यासाठी वापरली जाते.हे एक संख्या देते जी चलांमधील संबंधांची ताकद आणि दिशा दर्शवते.

कार्ल पियर्सनचे सहसंबंधाचे गुणांक दोन प्रकारे मोजले जाऊ शकते:

1. जेव्हा वास्तविक मध्यापासून विचलन घेतले जाते, तेव्हा सूत्र आहे-

$r = r(x, y) = \Sigma xy / \sqrt{\Sigma X^2 \Sigma Y^2}$

उदाहरण 2.7: खाली दिलेल्या डेटावरून विक्री आणि खर्च यांच्यातील गुणांक शोधा-

Roll No	Marks in Subject A	Marks in Subject B
1	48	45
2	35	20
3	17	40
4	23	25
5	47	45

उत्तर:

A विषयातील गुण X ने आणि B विषयातील गुण Y ने दर्शवू द्या.

X	(X-34)=x	X^2	Y	(Y-35)=y	Y^2	xy
48	14	196	45	10	100	140
35	1	1	20	-15	225	-15
17	-17	289	40	5	25	-85
23	-11	121	25	-10	100	110
47	13	169	45	10	100	130
ΣX=170	Σx=0	Σx^2=776	ΣY=175	Σy=0	Σy^2=550	Σxy=280

सरासरीची गणना केली जाऊ शकते $\frac{\Sigma X}{5}$

$$r = \frac{\Sigma xy}{\sqrt{\Sigma X^2 \Sigma Y^2}} = \frac{280}{\sqrt{776\ X\ 550}} = \frac{280}{653.3} = 0.429$$

r = 0.429 असल्याने याचा अर्थ A आणि B या दोन्ही विषयांमध्ये मध्यम सकारात्मक सहसंबंध आहे.

2. जेव्हा गृहीत मध्यापासून विचलन घेतले जाते, तेव्हा सूत्र आहे-

$r = r(x,y) = (N\ \Sigma dxdy - \Sigma dx\ \Sigma dy) / (\sqrt{N\ \Sigma dx^2 - (\Sigma dx)^2}\ \sqrt{N\ \Sigma dy^2 - (\Sigma dy)^2})$

"r" हा सहसंबंध गुणांक म्हणून ओळखला जातो.

उदाहरण 2.8: वरील समस्या (वरील उदाहरण :1) गृहीत धरून देखील सोडवता येते-विषय A मधील गुण X ने दर्शवू द्या आणि B विषयातील Y ने दर्शवू द्या आणि त्यातील विषय A आणि B गुणांचा मीन अनुक्रमे 35 आणि 40 गृहीत धरा-

X	(X-35) = dx	dx^2	Y	(Y-40) = dy	dy^2	dxdy
48	13	169	45	5	25	65
35	0	0	20	-20	400	0
17	-18	324	40	5	0	0
23	-12	144	25	-15	225	180
47	12	144	45	5	25	60
ΣX=170	Σdx=-5	Σdx^2=781	ΣY=175	Σdy=-25	Σdy^2=675	Σdxdy=305

$$r = \frac{(5X\ 305-9-5)9-25)}{\sqrt{(5X781-25)(5X675-625)}} = \frac{1525-125)}{\sqrt{(3905-25)(3375-625)}} = \frac{1400}{\sqrt{3880X2750}} = \frac{1400}{3266.50}$$

r=429

कोरिलेशन गुणांक: (Coefficient of Correlation): कार्ल पियर्सनच्या कोरिलेशनच्या गुणांकात खालील निरीक्षण गुणधर्म आहेत जसे-

1. r चे मूल्य नेहमी -1 ते +1 किंवा -1 ≤ r ≥ +1 दरम्यान असते.

2. सहसंबंधाची डिग्री r च्या मूल्याने व्यक्त केली जाते उदाहरणार्थ r चे मूल्य +1 व्हेरिएबल्स आहे उच्च सकारात्मक सहसंबंधित किंवा परिपूर्ण सकारात्मक सहसंबंध आहेत, जर ते -1 व्हेरिएबल्स जास्त असतील कारात्मक सहसंबंधित किंवा परिपूर्ण नकारात्मक सहसंबंध आणि जर ते 0 असेल तर त्यांच्यात कोणताही संबंध नाहीचल.
3. बदलाची दिशा चिन्ह (+ve) किंवा (-ve) द्वारे दर्शविली जाते.
4. हे दोन रिग्रेशन गुणांक $r = \sqrt{bxy * byx}$ चे भौमितीय मध्य म्हणून व्यक्त केले जाते.कार्ल पियर्सनच्या सहसंबंधाच्या गुणांकाच्या मर्यादा.

Karl Pearson's Coefficient of correlation & Spearman's rank Coefficient of correlation

कार्ल पियर्सनच्या सहसंबंधाच्या गुणांकाच्या काही मर्यादा आहेत- (Limitations of Karl Pearson's Coefficient of correlation)

1. हे केवळ चलांमधील रेखीय संबंध परिभाषित करते.
2. दोन व्हेरिएबल व्हॅल्यूजच्या चरम मूल्यांमुळे त्याचा अवाजवी परिणाम होतो.
3. पिअर्सनच्या गुणांकाची गणना काही वेळा त्रासदायक असते.

निर्धाराचे गुणांक (Coefficient of determination): निर्धाराचा गुणांक 'r^2' ने दर्शविला जातो '. आर2 एका व्हेरिएबलमधील भिन्नतेचे प्रमाण मोजते दुसऱ्याद्वारे स्पष्ट केले आहे. r आणि r^2 दोन्ही असोसिएशनचे सममित उपाय आहेत. दुसऱ्या शब्दांत, द X आणि Y चा सहसंबंध Y आणि X च्या सहसंबंधाप्रमाणे वेगळा असणार नाही. याचा अर्थ X दोन्ही व्हेरिएबल्समध्ये आहे आणि Y यापैकी जे डिपेंडंट व्हेरिएबल म्हणून घेतले आणि स्वतंत्र व्हेरिएबल म्हणून घेतले तर फरक पडत नाही, काय महत्त्वाचे म्हणजे त्या दोघांमधील संबंध.

निर्धाराची पदवी स्पष्ट केलेल्या भिन्नता आणि एकूण भिन्नतेचे गुणोत्तर म्हणून व्यक्त केली जाऊ शकते-

r^2 = स्पष्टीकरण केलेले भिन्नता / एकूण भिन्नता

r^2 चे कमाल मूल्य 1 आहे कारण Y मधील सर्व भिन्नता स्पष्ट करणे शक्य आहे परंतु ते नाही या सर्वांपेक्षा अधिक स्पष्ट करणे शक्य आहे.

निर्धाराच्या गुणांकाचा अर्थ (Interpreation of Coefficient of determination):

निर्धाराचा गुणांक टक्केवारीच्या दृष्टीने दोन चलांमधील संबंध मोजतो. हे आहे प्रतिसाद व्हेरिएबल्सच्या मूल्यामध्ये स्पष्ट केलेल्या भिन्नतेचे प्रमाण. उदाहरणार्थ जर r^2 चे मूल्य आहे 0.82 म्हणजे रिस्पॉन्स व्हेरिएबलमधील 82 टक्के फरक स्पष्टीकरणात्मक व्हेरिएबलद्वारे स्पष्ट करता येण्याजोगा आहे आणि ते बाकीच्या 18 टक्के फरकांबद्दल काहीही नमूद करत नाही जे इतर कोणत्याही घटकामुळे असू शकते किंवा चल

1. जर $r^2 = 0$ असे नमूद केले असेल तर दोन चलांमध्ये कोणताही संबंध नाही.
2. जर $r^2 = 1$ हा चलांमधील परिपूर्ण संबंध असेल.
3. जर $0 \le r^2 \le 1$ असेल तर प्रतिसाद व्हेरिएबलमधील भिन्नतेची डिग्री मूल्यांमधील फरकामुळे आहे स्पष्टीकरणात्मक चल. r^2 मूल्य शून्याच्या जवळ प्रतिसादातील फरकाचे कमी प्रमाण दाखवते स्पष्टीकरणात्मक व्हेरिएबलच्या मूल्यांमधील फरकामुळे चल. r^2 चे मूल्य असताना जवळपास 1 शो रिस्पॉन्स व्हेरिएबलमधील संपूर्ण फरक स्पष्टीकरणात्मक व्हेरिएबलच्या मूल्यांमधील फरकामुळे आहे

स्पिअरमॅनचा रँक कोरिलेशन गुणांक:(Spearman's rank Coefficient of correlation): जर, व्हेरिएबल्सचे परिमाणवाचकपणे मोजमाप करणे शक्य नसेल परंतु नंतर क्रमवारीत मांडले जाऊ शकते.कार्ल पियर्सनचा सहसंबंधाचा गुणांक लागू नाही. या परिस्थितींसाठी, च्या गुणांक ब्रिटीश मानसशास्त्रज्ञ चार्ल्स, एडवर्ड स्पिअरमन यांनी परस्परसंबंध दिले होते. या पद्धतीत द गटातील व्यक्ती अशा क्रमाने मांडल्या जातात ज्यायोगे प्रत्येक व्यक्तीला सूचित संख्या प्राप्त होते गटात त्याची रँक. या प्रकारच्या सहसंबंधांना सहसा नॉन-मेट्रिक

सहसंबंध म्हटले जाते कारण ते एक माप आहे दोन नॉन-मेट्रिक व्हेरिएबल्ससाठी सहसंबंध जे सहसंबंध मोजण्यासाठी रँकिंगवर अवलंबून असतात.

गुण (Merits):

- ही पद्धत अगदी सोपी आहे आणि तुलनेने संबंध प्रस्थापित करण्याचा सोपा मार्ग आहे
- स्पिअरमॅनचा रँक सहसंबंध गुणांक गैर-परिमाणवाचक किंवा गुणात्मक.
- या पद्धतीद्वारे दोघांमधील संबंध प्रस्थापित करण्यासाठी क्रमवार क्रमवारी सहज दिली जाऊ शकते गुणात्मक .

अवगुण (Demerits):

- असे गृहीत धरले जाते की दोन्ही व्हेरिएबल्स सामान्यपणे वितरीत केले जातात जे नेहमी खरे नसतात.
- हे केवळ व्हेरिएबल्समधील रेखीय संबंधांचे वर्णन करते, त्याबद्दल काहीही अर्थ लावत नाही नॉनलाइनर संबंध.
- मोठ्या डेटाच्या बाबतीत, रँक देणे थोडे कठीण आहे.
- गटबद्ध डेटा उपलब्ध असताना ही पद्धत लागू केली जाऊ शकत नाही.

स्पिअरमॅनच्या रँक कोरिलेशनची गणना (Calculation of Spearman's Rank Correlation):

सहसंबंधाचा रँक गुणांक तीन परिस्थितींनुसार मोजला जातो-

a) जेव्हा वास्तविक रँक दिले जातात तेव्हा सूत्र आहे-

$$R = 1 - \frac{(6\sum D^2)}{(N\,(N^2-1)}$$

जेथे R = रँक सहसंबंध गुणांक D = दोन मालिकेतील जोडलेल्या आयटममधील रँकमधील फरक. N= एकूण निरीक्षण संख्या.

गणना पद्धत (Calculation Method): जेथे वास्तविक रँक दिलेली आहेत तेथे या चरणांचे अनुसरण करून गणना केली जाऊ शकते-

- दोन श्रेणीतील फरक घ्या (R1-R2) हा फरक D द्वारे दर्शविला जातो.
- फरकांचे वर्गीकरण करून $\sum D^2$ ची गणना करा आणि त्यांची एकूण संख्या घ्या.
- सूत्र लागू करा-
- $R = 1 - \frac{(6\sum D^2)}{N\,(N^2-1)}$

उदाहरण 2.9: A आणि B या दोन विषयांतील दहा विद्यार्थ्यांचे गुण खालील तक्त्यामध्ये दिले आहेत-

Ranks in Subject A	Ranks in Subject B
3	4
5	6
4	3
8	9
9	10
7	77
1	2
2	1
6	5
10	8

स्पिअरमॅनचे रँक सहसंबंध सूत्र वापरून सहसंबंध गुणांक शोधा-

उत्तर:

पायरी 1: दोन रँकमधील फरक घ्या आणि त्याचे वर्ग करा-

Ranks in Subject A(R_1)	Ranks in Subject B(R_2)	$D^2 = (R_1 - R_2)^2$
3	4	1
5	6	1
4	3	1
8	9	1
9	10	1
7	77	0
1	2	1
2	1	1
6	5	1
10	8	4
		$\sum D^2 = 12$

पायरी 2: एकूण D^2 घ्या म्हणजे $\sum D^2 = 12$

पायरी 3: स्पिअरमॅनचे रँक गुणांक सूत्र वापरून गणना करा-

$$R = 1 - \frac{(6 \sum D^2)}{(N\,(N^2 - 1)}$$

$$= 1 - \frac{6 X 12}{(10^3 - 10)}$$

$$= 1 - \frac{72}{990}$$

$= 0.927$

b) जेव्हा रँक दिले जात नाहीत

रँक न दिल्यास उपलब्ध निरीक्षणांना रँक नियुक्त करणे आवश्यक आहे. आपण देणे सुरू करू शकतो.प्रथम किंवा शेवटचे ते सर्वोच्च किंवा सर्वात कमी देऊन रँक देतात परंतु आम्ही इतर सर्व उपलब्धांसाठी समान नियम पाळला पाहिजे $(R_1 - R_2)^2$

उदाहरण 2.10: दोन विषयांच्या खालील दिलेल्या गुणांसाठी रँक सहसंबंध गुणांकाची गणना करा A आणि B.

Marks of Subject A	92	89	87	86	83	77	71	63	53	50
Marks of Subject B	86	83	91	77	68	85	52	82	37	57

उत्तर: रँक सहसंबंध गुणांकाची गणना-

Marks of Subject A	R_1	Marks of Subject B	R_2	$(R_1 - R_2)^2 = D^2$
92	10	86	9	1
89	9	83	7	4
87	8	91	10	4
86	7	77	5	4
83	6	68	4	4
77	5	85	8	9
71	4	52	2	4
63	3	82	6	9
53	2	37	1	1
50	17	57	3	4
				$\sum D^2 = 44$

$$R = 1 - \frac{(6\sum D^2)}{(N\,(N^2-1)}$$

$$= 1 - \frac{6\,X\,44}{(10^3-10)}$$

$$= 1 - \frac{264}{990} = 0.733$$

क) जेव्हा रँक समान असतात

काही प्रकरणांमध्ये दोन किंवा अधिक डेटा समान असू शकतात म्हणून समान श्रेणी दिली जातात, त्या प्रकरणांमध्ये आम्ही नियुक्त करतो त्या प्रत्येकाला सरासरी रँक उदाहरणार्थ जर दोन व्यक्ती चौथ्या स्थानावर समान असतील तर प्रत्येकाला दिलेले सरासरी रँक $\frac{4+5}{2}$= 4.5 त्याचप्रमाणे चौथ्या स्थानावर तीन समान असल्यास, त्यांना सरासरी श्रेणी दिली जाते $\frac{4+5+6}{3}$= 5

स्पीयरमॅनच्या सहसंबंधाच्या गुणांकाची गणना करण्यासाठी समान श्रेणीचे सूत्र आहे –

$$R = 1 - \frac{6\left\{\sum D^2 + \frac{1}{12}(m_1{}^3 - m_1) + \frac{1}{12}(m_2{}^3 - m_2)\ldots\ldots\right\}}{N^3 - N}$$

जेथे m ही वस्तूंची संख्या आहे ज्यांची श्रेणी सामान्य आहे.

उदाहरण 2.11: आठ अर्जदारांचे X आणि Y या दोन विषयांतील गुण खाली दर्शविले आहेत. च्या रँक गुणांकाची गणना करा.

Applicants	A	B	C	D	E	F	G	H
Subject X	15	20	28	12	40	60	20	80
Subject Y	40	30	50	30	20	10	30	60

Applicants	**Marks in Subject X**	**Rank Assigned**	**Marks in Subject Y**	**Rank Assigned**	**$(R_1-R_2)^2=D^2$**
A	15	2	40	6	16
B	20	3.5	30	4	0.25
C	28	5	50	7	4
D	12	1	30	4	9
E	40	6	20	2	16
F	60	7	10	1	36
G	20	3.5	30	4	0.25
H	80	8	60	8	0
					$\sum D^2$=81.5

उत्तर:

$$R = 1 - \frac{6\left\{\sum D^2 + \frac{1}{12}(m_1{}^3 - m_1) + \frac{1}{12}(m_2{}^3 - m_2)\ldots\ldots\right\}}{N^3 - N}$$

विषय X मध्ये 20 दोन वेळा पुनरावृत्ती होते म्हणून m_1 = 2 आणि विषय Y 30 मध्ये तीन वेळा पुनरावृत्ती होते m_2= 3, ही मूल्ये सूत्रात टाकणे-

$$R = 1 - \frac{6\left\{81.5 + \frac{1}{12}(2^3-2) + \frac{1}{12}(3^3-3)\right\}}{(8^3-8)}$$

$$= 1 - \frac{6\,(81.5+0.5+2)}{(504)}$$

$= 1 - \frac{6 X 84}{504}$

= 0

R= 0 असल्याने याचा अर्थ दोन विषयांमध्ये मिळालेल्या गुणांमध्ये कोणताही संबंध नाही.

सारांश

- जेव्हा दोन व्हेरिएबल्स अशा प्रकारे बदलतात की कोणत्याही क्षणी एका व्हेरिएबलचे मूल्य शी संबंधित असते
- दुसऱ्या व्हेरिएबलच्या दोन्ही मूल्यांना सहसंबंधित चल म्हणतात. सहसंबंध विश्लेषण ही सांख्यिकीय पद्धत आहे
- दोन किंवा अधिक चलांमधील संबंधांची डिग्री मोजण्यासाठी. परस्परसंबंध विश्लेषण स्थापित करते
- व्हेरिएबल्समधील संबंध पण कोणते व्हेरिएबल हे कारण आहे आणि कोणते व्हेरिएबल आहे हे सांगत नाही
- परिणाम किंवा दुसऱ्या शब्दांत, चलांमधील कार्यकारण संबंध परस्परसंबंधाने स्थापित केला जाऊ शकत नाही

विश्लेषण

परस्परसंबंध वेगवेगळ्या प्रकारचे असू शकतात-

1. सकारात्मक किंवा नकारात्मक
2. साधे किंवा अनेक
3. रेखीय किंवा नॉनलाइनर

कोरिलेशनचा अभ्यास करण्याच्या विविध पद्धती आहेत, खाली नमूद केल्या आहेत-

1. स्कॅटर डायग्राम पद्धत
2. कार्ल पियर्सनचा कोरिलेशन गुणांक
3. स्पियरमॅनचा रँक कोरिलेशन

कोरिलेशन विश्लेषणामध्ये एक अतिशय महत्त्वाची संज्ञा ' कोरिलेशनचे गुणांक' मोजली जाते आणि द्वारे दर्शविली जाते'आर' 'r' चे मूल्य व्हेरिएबल्समधील संबंधांची ताकद परिभाषित करते. च्या गुणांकाचे मूल्यसहसंबंध 0 आणि 1 दरम्यान असू शकतो. r 1 च्या जवळ जेवढा असेल तितकाच चल संबंध अधिक मजबूत असेलr च्या संख्यात्मक मूल्याचे चिन्ह देखील संबंधांची सकारात्मक किंवा नकारात्मक दिशा दर्शवते. कोरिलेशनचा गुणांक देखील फक्त 'निर्धारित गुणांक' चे मूल्य निर्धारित करण्यासाठी वापरला जातोत्याचे वर्गीकरण करतो आणि r^2 द्वारे दर्शविला जातो. निर्धाराचे गुणांक r^2 एकामध्ये फरकाची टक्केवारी सांगते.

2.7 Regression equation of line in two variable

रिग्रेशनचा अर्थ आणि प्रकार (Meaning and Types of Regression)

- कमीत कमी चौरस पद्धत (Least Square Method)

- X वर Y चे रिग्रेशन (Regression of Y on X)

- Y वर X चे रिग्रेशन (Regression of X on Y)

रिग्रेशन गुणांकाचे गुणधर्म (Properties of Regression Coefficient)

रिग्रेशनचा अर्थ आणि प्रकार (Meaning and Types of Regression):

- **रिग्रेशनचा अर्थ(Meaning of Regression)**

रेखीय रिग्रेशन ही एका व्हेरिएबलच्या मूल्याचा अंदाज लावण्याची एक पद्धत आहे जेव्हा इतर सर्व चलांची मूल्ये ज्ञात किंवा निर्दिष्ट केली जातात.ज्या व्हेरिएबलचा अंदाज लावला जातो त्याला रिस्पॉन्स किंवा डिपेंडेंट व्हेरिएबल म्हणतात. प्रतिसादाचा अंदाज लावण्यासाठी वापरल्या जाणाऱ्या चलना किंवा अवलंबित चलना प्रेडिक्टर किंवा स्वतंत्र चल म्हणतात. रेखीय रिग्रेशन हे प्रस्तावित करते की रेखीय समीकरणाद्वारे वर्णन केलेल्या दोन किंवा

अधिक चलांमधील संबंध. या उद्देशासाठी वापरल्या जाणाऱ्या रेखीय समीकरणाला रेखीय आवर्तन मॉडेल म्हणतात. रेखीय रिग्रेशन मॉडेलमध्ये अज्ञात गुणांक असलेले रेखीय समीकरण असते. रेखीय रिग्रेशन मॉडेलमधील अज्ञात गुणांकांना रेखीय रिग्रेशन मॉडेलचे पॅरामीटर्स म्हणतात. मॉडेलच्या अज्ञात पॅरामीटर्सचा अंदाज घेण्यासाठी व्हेरिएबल्सची निरीक्षण केलेली मूल्ये वापरली जातात. उपलब्ध नमुना डेटा वापरून दोन किंवा अधिक चलांमधील संबंध दर्शवण्यासाठी एक रेखीय समीकरण विकसित करण्याची प्रक्रिया म्हणून ओळखली जाते

- **रेखीय रिग्रेशनचे प्रकार(Types of Linear Regression)**

रेखीय रीग्रेशनचे प्राथमिक उद्दिष्ट दोन किंवा अधिक चलांमधील संबंध व्यक्त करण्यासाठी किंवा प्रतिनिधित्व करण्यासाठी एक रेखीय समीकरण विकसित करणे आहे. रिग्रेशन समीकरण हे गणितीय समीकरण आहे .

- **समर्पक साधे रेखीय रिग्रेशन (Fitting Simple Linear Regression)**

एक उदाहरण विचारात घ्या जिथे आपल्याला पाहिजे आहे पिकाच्या उत्पन्नाचा अंदाज लावा (किलो. प्रति एकर)खताच्या प्रमाणात एक रेषीय कार्य म्हणून लागू (किलो. प्रति एकर). या उदाहरणात, द पीक उत्पन्नाचा अंदाज लावायचा आहे. म्हणून, ते आहे अवलंबून व्हेरिएबल आणि Y. द द्वारे दर्शविले जाते लागू केलेल्या खताची मात्रा बदलते अंदाज बांधण्याच्या उद्देशाने वापरले जाते.म्हणून, हे स्वतंत्र चल आहे आणि आहे X द्वारे दर्शविले जाते.

Table2.1

खत ची किंमत(X)(किलो. प्रति एकर) (Fertilizer)	उत्पन्न (Y) (00 किलो मध्ये) (Yield)
30	43
40	45
50	54
60	53
70	56
80	63

Table 2.1 खताची मात्रा दर्शवते आणि सहा प्रकरणांसाठी पीक उत्पन्न. या जोड्या निरिक्षणांचा वापर स्कॅटर मिळविण्यासाठी केला जातो
Fig 2.1 मध्ये दर्शविल्याप्रमाणे आकृती (Scatter Diagram)

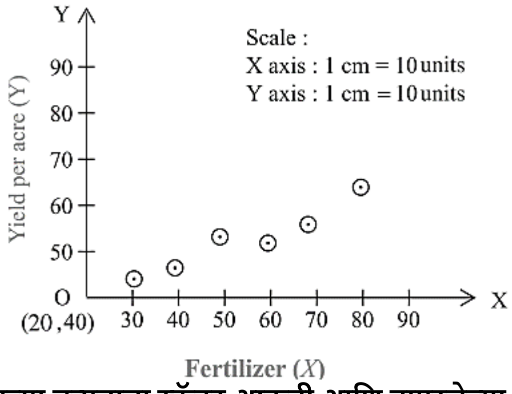


Fig 2.1: धान्याच्या उत्पन्नाचा स्कॅटर आकृती आणि वापरलेल्या खताची मात्रा.

आम्हाला सरळ रेषा काढायची आहे स्कॅटर डायग्राममधील बिंदूंच्या सर्वात जवळ (Fig 3.1). जर सर्व बिंदू समरेखित असतील (ते आहे, सरळ रेषेवर), तेथे असती अशी रेषा काढायला हरकत नाही. आहे एक समस्या

कारण सर्व बिंदू a वर नाहीत सरळ रेषा.स्कॅटर डायग्राममधील बिंदू हे करतात notforma सरळ रेषा, आम्हाला सरळ काढायची आहे रेषा जी या बिंदूंच्या सर्वात जवळ आहे. सैद्धांतिकदृष्ट्या,संभाव्य रेषेची संख्या अमर्यादित आहे. हे आहे म्हणून काही अट निर्दिष्ट करणे आवश्यक आहे आम्ही सरळ रेषा काढतो याची खात्री करण्यासाठी ते स्कॅटरमधील सर्व डेटा पॉइंट्सच्या सर्वात जवळ आहे आकृती कमीतकमी चौरसांची पद्धत प्रदान करते सर्वोत्तम फिटची ओळ कारण ती डेटाच्या सर्वात जवळ आहे नुसार स्कॅटर डायग्राममधील बिंदू किमान चौरस तत्त्व.

- **सर्वात कमी चौरसांची पद्धत (Method of Least Square):** सर्वोत्कृष्ट फिटची ओळ मिळविण्यासाठी वापरलेले तत्व ला किमान चौरसांची पद्धत म्हणतात.कमीत कमी चौरसांची पद्धत विकसित झाली अॅड्रिन-मायरे लॅजेंड्रे आणि कार्ल फ्रेडरिक यांनी गॉस एकमेकांपासून स्वतंत्रपणे. चला तत्त्वामागील मध्यवर्ती कल्पना समजून घ्या किमान चौरसांची. समजा डेटामध्ये n च्या जोड्यांचा समावेश आहे मूल्ये (x1 , y1), (x2, y2), ..., (xn, yn) आणि समजा की दिलेल्या डेटाशी उत्तम बसणारी ओळ आहे खालीलप्रमाणे लिहिले आहे. Ŷ =a + bX (येथे, Ŷ आहे Y कॅप म्हणून वाचिचे.) या समीकरणाला म्हणतात अंदाज समीकरण. तेच वापरत आहे स्थिरांकांची मूल्ये a आणि b, अंदाजित मूल्य Y चे Ŷi =a + bXi द्वारे दिलेले आहेत , जेथे xi आहे स्वतंत्र व्हेरिएबलचे मूल्य आणि Yi आहे Y चे संबंधित अंदाजित मूल्य. लक्षात ठेवा स्वतंत्र व्हेरिएबलचे निरीक्षण केलेले मूल्य yi Y हे अंदाजित मूल्यापेक्षा वेगळे आहे Yi .

निरीक्षण मूल्यवान (yi) आणि अंदाज मूल्ये Yi Y पैकी ( ) पूर्णपणे जुळत नाही कारण निरीक्षणे सरळ रेषेवर पडत नाहीत. दनिरीक्षण मूल्ये आणि मध्ये फरक अंदाजित मूल्यांना त्रुटी किंवा अवशेष म्हणतात.^Ŷ गणिती बोलणे प्रमाण $y_1 - Y_1, y_2 - Y_2$ -------- $y_n - Y_n$ किंवा समतुल्य,

प्रमाण $y_1 - (a + bx_1)$, $y_2 - (a + bx_2)$, ..., $y_n - (a + bx_n)$ हे निरीक्षण केलेल्या मूल्यांचे विचलन आहे संबंधित अंदाजित मूल्यांमधून Y चा आणि म्हणून त्यांना त्रुटी किंवा अवशेष म्हणतात. आम्ही लिहा ei= yi – Y1∝ = yi - (a + bxi), i = 1, 2, ..., n साठी. भौमितिकदृष्ट्या, अवशिष्ट ei, जे yi – (a + bxi), अनुलंब सूचित करते. अंतर (जे सकारात्मक किंवा नकारात्मक असू शकते) निरीक्षण मूल्य दरम्यान (yi) आणि अंदाज मूल्य (Yi).

किमान वर्गाच्या पद्धतीचे तत्त्व खालील प्रमाणे सांगितले जाऊ शकते.

सर्व शक्य सरळ रेषांमध्ये की स्कॅटर आकृतीवर, ची रेषा काढली जाऊ शकते. सर्वोत्कृष्ट तंदुरुस्त ची व्याख्या अशी आहे जी कमी करते. अवशेषांच्या वर्गांची बेरीज, म्हणजेच बेरीज अंदाजित y-मूल्यांच्या विचलनांचे वर्ग निरीक्षण केलेल्या y-मूल्यांमधून. दुसऱ्या शब्दांत, द सर्वोत्तम फिटची ओळ निर्धारित करून प्राप्त केली जाते स्थिर a आणि b जेणेकरून किमान आहे.

$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}[y_i - (a + b\boldsymbol{x_i})]^2$

हे वापरून प्राप्त केलेली सरळ रेषा तत्त्वाला सर्वात कमी रिग्रेशन रेषा म्हणतात. लाक्षणिकरित्या, आम्ही लिहितो

$$S^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

चौरस त्रुटींची बेरीज म्हणून**.** हे देखील असू शकते म्हणून लिहिले

$$S^2 = \sum_{i=1}^{n}[y_i - (a + b\boldsymbol{x_i})]^2$$

आम्हाला a आणि b स्थिरांक ठरवायचे आहेत अशा प्रकारे की S2 किमान आहे.लक्षात घ्या की S2 एक सतत आणि आहे a आणि b दोन्हीचे भिन्न कार्य. आम्ही a च्या संदर्भात S2 मध्ये फरक करा (ब ते गृहीत धरून स्थिर रहा) आणि b च्या संदर्भात (अ गृहीत धरून स्थिर) आणि b च्या संदर्भात (a आहे असे गृहीत धरून स्थिर).

त्यानंतर आम्ही या दोन्ही व्युत्पन्नांची समानता करतो S2 कमी करण्यासाठी शून्यावर. परिणामी,आपल्याला खालील दोन रेषीय समीकरणे दोन मध्ये मिळतात अज्ञात a आणि b.

$\sum_{i=1}^{n} y_i$ **=na+**$b\sum_{i=1}^{n} x_i$

$\sum_{i=1}^{n} x_i y_i$ **=** $a\sum_{i=1}^{n} x_i$ **+**$b\sum_{i=1}^{n} {x_i}^2$

जेव्हा आपण ही दोन रेखीय समीकरणे सोडवतो**, a** आणि **b** ची मूल्ये कमी करतात जी **S2**दिले द्वारे जातात

a=$\overline{y}$ **-b**$\overline{x}$

b = $\frac{cov\ (x,y)}{{\sigma_x}^2}$**,**

cov(x.y) = $\frac{1}{n}\sum_{i=1}^{n} x_i y_i$ **-**$\overline{xy}$ **and**

${\sigma_x}^2$**=** $\frac{1}{n}\sum_{i=1}^{n} {x_i}^2 - \overline{x}^2$

प्राप्त केलेल्या a आणि b ची मूल्ये बदलणे रिग्रेशन समीकरणात वर दर्शविल्याप्रमाणे

$$Y = a + bX,$$

आम्हाला समीकरण मिळते

$Y - \overline{y}$ =b (X-$\overline{x}$)

X वर Y चे रिग्रेशन (Regression of Y on X)

आम्ही आता b साठी b_{YX} नोटेशन सादर करतो.जेव्हा Y हे अवलंबून चल असते आणि X असते स्वतंत्र चल. X वर Y चे रेखीय रिग्रेशन असे गृहीत धरते व्हेरिएबल X हे स्वतंत्र व्हेरिएबल आहे आणि व्हेरिएबल Y हे अवलंबून व्हेरिएबल आहे. क्रमाने हे स्पष्ट करण्यासाठी, आम्ही रेखीय व्यक्त करतो खालीलप्रमाणे रिग्रेशन मॉडेल.

$Y - \overline{y}$ =b_{xy} (X-$\overline{x}$)

or

Y=b_{xy}X

येथे, b ची जागा b_{YX} ने घेतली आहे हे लक्षात घ्या

$b_{yx}= \frac{cov(X,Y)}{Var(X)}$

$= \frac{\sum(x_i-\overline{x})(y_i-\overline{y})}{\sum(x_i-\overline{x})^2}$

$= \frac{\sum x_i y_i - n\overline{xy}}{\sum {x_i}^2 - n\overline{(x)}^2}$

Y वर X चे रिग्रेशन. (Regression of X on Y)

नोटेशन $b_{\mathbf{XY}}$ म्हणजे b जेव्हा X असतो अवलंबून चल आणि Y स्वतंत्र आहे चल

Y वर X चे रेखीय रिग्रेशन असे गृहीत धरते व्हेरिएबल Y हे स्वतंत्र व्हेरिएबल आहे आणि व्हेरिएबल X हे अवलंबून व्हेरिएबल आहे. क्रमाने हे मॉडेल वेगळे आहे हे स्पष्ट करण्यासाठी X वर Y चे रेखीय रिग्रेशन, आपण व्यक्त करतो खालीलप्रमाणे रेखीय रिग्रेशन मॉडेल.

$$X = a' + b'Y.$$

किमान चौरसांची पद्धत, लागू केल्यावर या मॉडेलमध्ये खालील अभिव्यक्ती होतात स्थिर a′ आणि b′ साठी.

$a' = \overline{x} - b'\overline{y}$ आणि

$b' = \frac{cov(X,Y)}{Var(X)}$

मध्ये a′ आणि b′ ची मूल्ये बदलणे रेखीय रिग्रेशन मॉडेल, आम्हाला मिळते

X=a'+b'y

$X = \overline{x} - b'\overline{y} + b'y$

$X - \overline{\boldsymbol{x}}$ =b (Y-$\overline{y}$)

टीप: वरील समीकरणातील स्थिर b′Y. In वर X चा रिग्रेशन गुणांक म्हणतात हे स्पष्ट करण्यासाठी, यापुढे ते होईल.

b′ ऐवजी bXY असे लिहिले आहे. सर्वात कमी चौरस म्हणून Y वर X चे रिग्रेशन असे लिहिले जाईल.

$(X - \overline{\boldsymbol{x}})$ =b YX (Y-$\overline{y}$)

लक्षात घ्या की Y वर X चे रेखीय रिग्रेशन आहे म्हणून व्यक्त केले.

येथे लक्षात घ्या की b च्या जागी b_{XY} ने घेतले आहे. हे करू शकता म्हणून लिहावे

$b_{xy} = \frac{cov(X,Y)}{Var(Y)}$

$= \frac{\frac{1}{n}\Sigma(x_i - \overline{x})(y_i - \overline{y})}{\frac{1}{n}\Sigma(y_i - \overline{y})^2}$

दाखवत आहे की स्थिरांक 1

$$= \frac{\frac{1}{n}\sum x_i y_i - \overline{x}\,\overline{y}}{\frac{1}{n}\sum y_i^2 - \overline{y}^2}$$

$$b_{XY} = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sum(y_i - \overline{y})^2}$$

$$= \frac{\sum x_i y_i - n\overline{x}\,\overline{y}}{\sum x_i^2 - n(\overline{x})^2}$$

उदाहरण 2.12: खत वापराच्या डेटासाठी आणिधान्याचे उत्पन्न तक्ता 3.2 मध्ये दिले आहे.

1. उत्पन्नाच्या रिग्रेशनची रेषा मिळवा खताच्या प्रमाणात धान्य वापरले.
2. कमीत कमी चौरस रिग्रेशन रेषा काढा.
3. धान्याच्या उत्पन्नाचा अंदाज केव्हा 90 किलो च्या खताचा वापर केला जातो.

उत्तर:

खतची किंमत X = xi	उत्पन्न Y = yi	Xi^2	xiyi
30	43	900	1290
40	45	1600	1800
50	54	2500	2700
60	53	3600	3180
70	56	4900	3920
80	63	6400	5040
330	314	19900	17930

उत्तर:

Since n=6 $\sum_{i=1}^{6} xi = 330$ $\sum_{i=1}^{6} yi = 314$

$$\sum_{i=1}^{n} xi^2 = 19900 \quad \sum_{i=1}^{n} xiyi = 17930$$

$\overline{x} = \sum \frac{x_i}{n} = 55$ $\overline{y} = \sum \frac{y_i}{n} = 52.3$

$b_{xy} = \frac{\sum x_i y_{i-n\overline{x}\overline{y}}}{\sum {x_i}^2 - n\overline{x}^2}$

$= \frac{17930-6X55X52.3}{19900-6X3025}$

$= \frac{671}{1750} = 0.38$

$\mathbf{a} = \overline{y} - b_{xy} . \overline{x}$

.शेवटी x वरील y च्या रिग्रेशनची रेषा आहे

Y=31.4+ 0.38 X

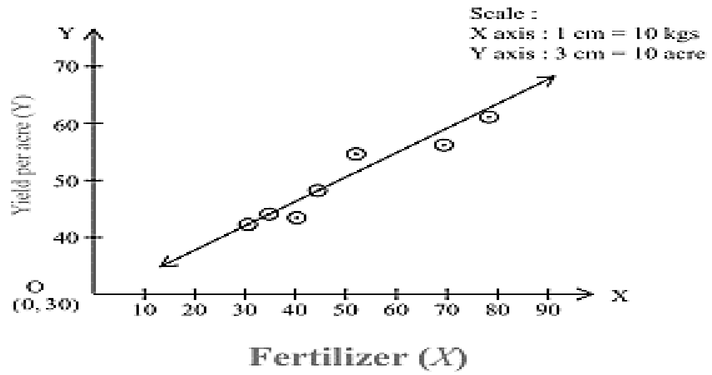


Fig. 3.2 : Scatter diagram and the line of regression of yield on amount of fertilizer

2. कमीत कमी चौरस रिग्रेशन रेषा काढण्यासाठी, आम्ही X ची कोणतीही दोन सोयीस्कर मूल्ये निवडतो आणि Y ची संबंधित मूल्ये शोधा

x = 35. y = 44.7 साठी

x = 45, y = 48.5

दोन गुण जोडणे (35,44.8) आणि (45,48.6), आम्हाला Fig 3.2 मध्ये ओळ मिळते

3. रिग्रेशन समीकरणात x = 90 टाकणे

Ŷ= 31.4 + 0.38 × 90 = 65.6

उदाहरण 2.13: डिपार्टमेंटल स्टोअर सेल्समनना प्रशिक्षण आणि त्यानंतर चाचणी सेवेत देते. कोणत्याही सेल्समनची संबंधित कामगिरीचा अनुभव ते त्याने मिळवलेले विक्री गुण रेखीयरित्या संबंधित आहे. खालील डेटा चाचणी गुण आणि नऊ सेल्समननी केलेली विक्री निश्चित कालावधी देतो.

चाचणी गुण (X)	16	22	28	24	29	25	16	23	24
विक्री ('00 Rs.) (Y)	35	42	57	40	54	51	34	47	45

1. X वर Y ची रिग्रेशन रेषा मिळवा.
2. X = 17 असताना Y चा अंदाज लावा.

उत्तर: गणिते स्पष्टपणे दर्शविण्यासाठी, खालील तक्ता तयार आहे

X=xi	Y=yi	xi− x	yi−y	(xi− x)^2	(xi− x)(yi−y)
16	35	-7	-10	49	70
22	42	-1	-3	1	3
28	57	5	12	25	60
24	40	1	-5	1	-5

29	54	6	9	36	54
25	51	2	6	4	12
16	34	-7	-11	49	77
23	47	0	2	00	00
24	45	1	0	1	00
207	405	00	00	166	271

n=9 and $\sum_{i=1}^{9} xi = 207$ $\sum_{i=1}^{9} yi = 405$

$\overline{x} = \sum \frac{x_i}{n} = 23$ $\overline{y} = \sum \frac{y_i}{n} = 45$

कारण X आणि Y ची साधने संपूर्ण आहेत संख्या, सूत्र वापरणे श्रेयस्कर आहे

$$\frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sum(x_i - \overline{x})^2}$$

for calculation of byx

Line of regression of Y on X

Y=a+b_{xy} X

$b_{xy} = \frac{Cov(X,Y)}{\sigma_x{}^2}$

$= \frac{\frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{n}}{\frac{\sum(x_i - \overline{x})^2}{n}}$

$= \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sum(x_i - \overline{x})^2}$

$= \frac{271}{166} = 1.6325$

And **a** = $\boldsymbol{\overline{y}}$ **-**$\boldsymbol{b_{xy}}$ **.** $\boldsymbol{\overline{x}}$ **= 7.4525**

येथे, X वर Y च्या रिग्रेशन रेषा आहे(Line of Regression of Y on X)

Y = 7.4525 + 1.6325X

ii) X = 17 असताना Y चा अंदाज(Estimate of Y when X=17)

Y = 7.4525 + (1.6325) (17)

= 35.205

उदाहरण 2.14: मोठ्या फर्निचर स्टोअरचे व्यवस्थापन विक्री(हजारो मध्ये रु.) (X) संख्येच्या आधारावर दिलेल्या दिवशी त्या दिवशी स्टोअरला भेट दिलेल्या लोकांची (Y) संख्या निश्चित करू इच्छितो . आवश्यक नोंदी ठेवल्या गेल्या, आणि एक याद्दच्छिक दहा दिवसांचा नमुना अभ्यासासाठी निवडण्यात आला.सारांश परिणाम खालीलप्रमाणे होते .

$\sum x_i = 370, \sum y_i = 580$ $\sum x_i{}^2 = 17200$

$\sum y_i{}^2 = 41640, \sum x_i y_i = 11500$ *,n=10*

x वरील y च्या रिग्रेशनची रेषा शोधा.

उत्तर: x वरील y च्या रिग्रेशनची रेषा

X=a'+b_{xy}Y

$b_{xy} = \frac{Cov(X,Y)}{\sigma_y{}^2}$ $\overline{x} = \sum \frac{x_i}{n} = 37$ $\overline{y} = \sum \frac{y_i}{n} = 58$

$$b_{xy} = \frac{\frac{\sum x_i y_i}{n} - \overline{xy}}{\frac{\sum y_i{}^2}{n} - \overline{y^2}}$$

$$b_{xy}= \frac{\frac{11500}{10}-(\frac{370}{10})(\frac{580}{10})}{\left[(\frac{41640}{10})(\frac{580}{10})\right]^2}$$

$$=\frac{996}{800}$$

$$= 1.245$$

$\overline{x}$ -b_{xy} . $\overline{y}$

= 37- (-1.245)(58)= 109.21

x वरील y च्या रिग्रेशनची रेषा

X=109.21 - 1.245Y

सराव उदाहरणे

1. एखाद्या कंपनीच्या मनुष्यबळ विकास व्यवस्थापकाला हवे आहे कि उत्पादन विभागात नोकरी अर्ज करणाऱ्या व्यक्तींचे मासिक उत्पन्न निराकरण करण्यासाठी एक उपाय शोधा. म्हणून त्यांनी त्या विभागातील व्यक्तींचा वर्षांची सेवा आणि त्यांचे मासिक उत्पन्न 7 चा डेटा गोळा केला.

व्यक्तींचा वर्षांची सेवा	11	7	9	5	8	6	10
सेवेचे (X) मासिक उत्पन्न (Rs. 1000) (Y)	10	8	6	5	9	7	11

वर्षांच्या सेवेवर उत्पन्नचे रिग्रेशन समीकरण शोधा.

2. खालील डेटावरून X वर Y आणि Y वर X च्या रिग्रेशन समीकरणांची गणना करा

X	10	12	13	17	18
Y	5	6	7	9	13

3. पाच (5) जोड्यांवर एका विशिष्ट द्विवेरिएट डेटासाठी निरीक्षणे दिली

$\sum x = 20, \sum y = 20, \sum x^2 = 90,$

$\sum y^2 = 90, \sum xy = 76$

गणना करा (i) cov(x,y) (ii) b_{yx} and b_{xy} (iii) r

4. खालील डेटावरून x = 125 असताना y ची किंमत काढा

X	120	115	120	125	126	123
Y	13	15	14	13	12	14

5. खालील तक्ता याद्दच्छिकपणे 10 निवडलेले कामगार चाचणी गुण आणि उत्पादकता निर्देशांक योग्यता देतो.

चाचणी गुण (X)	60	62	65	70	72	48	53	73	65	82
उत्पादकता निर्देशांक (Y)	68	60	62	80	85	40	52	62	60	81

दोन रिग्रेशन समीकरणे मिळवा आणि

a. ज्याचा चाचणी गुण 95 आहे त्या कामगाराचा उत्पादकता निर्देशांक काढा.

b. जेव्हा उत्पादकता निर्देशांक 75 आहे त्या कामगाराची चाचणी गुण काढा.

6. खालील डेटासाठी योग्य रिग्रेशनची गणना करा:

X (Independent variable)	2	4	5	6	8	11
Y (Dependent variable)	18	12	10	8	7	5

7. अर्थशास्त्रातील विद्यार्थ्यांद्वारे (X) आणिगणित (Y) मिळालेले गुण खालीलप्रमाणे आहेत

X	59	60	61	62	63
Y	78	82	82	79	81

X वर Y चे रिग्रेशन समीकरण शोधा.

8. खालील द्विवैरिएट डेटासाठी दोन रिग्रेशन समीकरणे प्राप्त करा:

X	1	2	3	4	5
Y	5	7	9	11	13

9. खालील डेटावरून समीकरण दोन रिग्रेशनओळींचे मिळवा:

X	6	2	10	4	8
Y	9	11	5	8	7

10. खालील डेटासाठी रिग्रेशन X वर Y ची ओळ शोधा

X	1	2	3
Y	2	1	6

x = 4 तेव्हा y चे संभाव्य मूल्य शोधा.

11. खालील डेटावरून, रिग्रेशन X वर Y चे समीकरण आणि X = 10 असताना Y ची किंमत शोधा

X	1	2	3	4	5	6
Y	2	4	7	6	5	6

12. खालील नमुना 12 विद्यार्थीच्या दररोज अभ्यासाचे तास (X) आणि गुण (Y) मिळालेले क्रमांक देतो .

X	3	3	3	4	4	5	5	5	6	6	7	8
Y	45	60	55	60	75	70	80	75	90	80	75	85

अभ्यासाचे तास वर मार्क्सची रिग्रेशन लाइन मिळवा.

संदर्भ (References):

1. http://www.purplemath.com/
2. https://www.khanacademy.org/
3. https://archive.nptel.ac.in/courses/111/105/111105077/

युनिट - 3
प्रोबॅबिलिटी ऑफ रँडम व्हेरिएबल्स
(Probability of Random Variable)

अभ्यासक्रम निष्पत्ती (Course Outcome):

CO3: दिलेल्या समस्यांचे निराकरण करण्यासाठी संभाव्यतेची तत्त्वे वापरा

घटक निष्पत्ती (Theory Learning Outcome):

1. बेरीज आणि गुणाकार संभाव्यता प्रमेय वापरून समस्या सोडवा.
2. सशर्त संभाव्यता वापरून समस्या सोडवा
3. बाईस' सिद्धांत वापरून समस्या सोडवा.

संभाव्यता (Probability)

3.1 संभाव्यता - व्याख्या (Definition)

3.2 संभाव्यतेचे बेरीज आणि गुणाकाराचे प्रमेय
(Theroem of Probability - Addition and Multiplication)

3.3 सशर्त संभाव्यता (Conditional Probability)

3.4 बाईस' सिद्धांत (Bayes ' Theorem)

ओळख (INTRODUCTION)

खालील विधाने काळजीपूर्वक वाचा

i. **बहुतेक** आज पासून पाऊस पडेल.
ii. महागाई वाढण्याचा **संभव** खूप आहे.
iii. **निश्चितच** मला प्रथम श्रेणी मिळणार.

बहुतेक , संभव , निश्चित यासारखे शब्द शक्यता दर्शवतात. थोडक्यात दैनंदिन जीवनात (day -to-day life) शक्यतांचा विचार आपण करत असतो.

संभाव्यता हा शब्द अनिश्चिततेच्या घटनेशी संबंधित आहे. संभाव्यता सिद्धांत म्हणजे अनिश्चिततेचा अभ्यास.

3.1 व्याख्या

3.1.1 प्रयोग (Experiment): ज्या कृती किंवा कामगिरीची पुनरावृत्ती कमी-अधिक समान स्थितीत केली जाऊ शकते आणि त्याचे काही विशिष्ट परिणाम असतील त्याला प्रयोग (Experiment) असे म्हणतात.

उदा.

1. नाणे हवेत फेकणे आणि छापा (H) किंवा काटा (T)आहे की नाही हे तपासणे याला प्रयोग (Experiment) म्हणतात.
2. हायड्रोजन (H_2) आणि ऑक्सिजन (O_2) यांची रासायनिक अभिक्रिया करून पाणी मिळवणे हा प्रयोग (Experiment) आहे.
3. 52 प्लेइंग कार्डमधून कार्ड काढणे आणि कार्डचा प्रकार तपासणे हा देखील प्रयोग आहे.

प्रयोगाचे दोन प्रकारचे आहेत.

1. **यादृच्छिक प्रयोग (Random Experiment)**: एखाद्या प्रयोगाचे परिणाम समान परिस्थितीत पुनरावृत्ती होत असतानाही सारखे नसतील तर त्याला यादृच्छिक प्रयोग (Random Experiment) म्हणतात. उदा. नाणे फेकणे, फासे टाकणे हे यादृच्छिक प्रयोगाचे साधारण उदाहरण आहेत.
2. **निर्धारवादी प्रयोग (Deterministic Experiment):**जर प्रयोगाचे परिणाम त्याच प्रारंभिक स्थितीत पुनरावृत्ती केल्यावर बदलत नाहीत तर त्याला निर्धारात्मक प्रयोग (Deterministic Experiment)

म्हणतात. उदा. हायड्रोजन (H_2) आणि ऑक्सिजन (O_2) पासून पाणी (Water – H_2O) तयार करणे.

3.1.2 रँडम एक्सपेरिमेंट-यादृच्छिक प्रयोग (Random Experiment): या प्रयोगात सर्व संभाव्य फलिते अगोदरच माहित असतात. पण त्यापैकी कोणत्याही फलित बद्दल निश्चित भाकीत आपण करू शकत नाही . सर्व फलिते सत्य असण्याची शक्यता समान असते अशा प्रयोगाला यादृच्छिक प्रयोग (Random Experiment) असे म्हणतात. उदा. फासा फेकणे आणि समोर दिसणाऱ्या नंबरचे निरीक्षण करणे हा यादृच्छिक प्रयोग आहे कारण समोर दिसणारा क्रमांक {1, 2, 3, 4, 5, 6} या संचातील असू शकतो, ज्याचा प्रयोग करण्यापूर्वी अंदाज लावता येत नाही.

वेगवेगळ्या रंगांचे बॉल असलेल्या बॉक्समधून बॉल घेणे आणि विशिष्ट रंगाच्या बॉलचे निरीक्षण करणे हा एक यादृच्छिक प्रयोग आहे. बल्बच्या गटातून बल्बच्या आयुष्याची लांबी तपासणे हा एक यादृच्छिक प्रयोग आहे. सर्व प्रकारच्या संशोधन उपक्रमांमध्ये यादृच्छिक प्रयोग सामान्य आहेत.

उदाहरणार्थ,

1. लोकांच्या गटावर नवीन औषध देणे आणि त्याची प्रभावीता पाहणे.
2. कारखान्यातून उत्पादन घेणे आणि ते सदोष आहे की नाही याची पडताळणी करणे.
3. उत्तम शैक्षणिक कामगिरीसाठी प्रशिक्षण देणे आणि त्याची परिणामकारकता मोजणे.

3.1.3 निष्पत्ती (Outcomes): यादृच्छिक प्रयोगाच्या फलिताला निष्पत्ती (Outcomes) म्हणतात

उदा.

1. एक नाणे फेकणे या प्रयोगाच्या दोनच निष्पत्ती आहेत. छापा (H) किंवा काटा (T)
2. एक फासा टाकणे या यादृच्छिक प्रयोगात त्याच्या 6 पृष्टभागावर असणाऱ्या बिंदूंच्या संख्येवरून 6 निष्पत्ती शक्य आहेत. 1 किंवा 2 किंवा 3 किंवा 4 किंवा 5 किंवा 6.

3.1.4 नमुना अवकाश (Sample Space): यादृच्छिक प्रयोगाच्या सर्व साध्या परिणामांच्या किंवा नमुना बिंदूंच्या संचाला नमुना अवकाश (Sample space) म्हणतात. नमुना अवकाश (Sample space) हे ‘S’ आणि ‘Ω’(Omega) अक्षराने दर्शविले जाते .

उदा. 1) एक नाणे फेकणे S={H , T } n(S)= 2

2) दोन नाणे फेकणे S={HH ,HT, TH, TT } n(S)= 4

3) तीन नाणे फेकणे S={HHH , HHT, HTH, TTH, THT, HTT, THH ,TTT } n(S)= 8

4) एक फासा टाकला S={1, 2, 3, 4, 5, 6 } n(S)= 6

3.1.5 घटना (Events): विशिष्ट अट पूर्ण करणाऱ्या निष्पत्तीला अपेक्षित निष्पत्ती (favourable outcome) म्हणतात. नमुना अवकाश दिला असेल तर अपेक्षित निष्पत्तीच्या संचाला 'घटना ' (Events) म्हणतात. घटना हा नमुना अवकाशाचा (Sample space) उपसंच असतो. अशाप्रकारे घटना हे नमुना जागेचे उपसंच आहेत ज्यात संभाव्यता विचाराधीन फंकशन्स (functions) द्वारे संभाव्यता नियुक्त केल्या जातात.

3.1.6 घटना चे प्रकार (Types of Events)

1) एकघटकी घटना (Simple events) :

ज्या घटनेला फक्त एक साधा घटक असतो त्याला एकघटकी घटना (Simple events) म्हणतात.

उदा. नाणे फेकण्याच्या प्रयोगात छापा (H) मिळण्याची घटना.

2) संयुक्त घटना (Compound events): एकापेक्षा जास्त घटक असलेल्या घटनेला संयुक्त घटना (Compound events) म्हणतात. उदा. एक फासा टाकल्यानंतर सम (even) किंवा विषम (odd) संख्या येणे.

3) अशक्य घटना (Impossible Events): प्रयोगात कोणताही नमुना बिंदू रिकाम्या संचाचा (Empty set denoted by Ø- phi) नसतो आणि त्यामुळे घटना कधीही होऊ शकत नाही. अशा प्रकारे रिकाम्या संच घटनेला अशक्य घटना (Impossible Events) म्हणतात. त्याची संभाव्यता (Probability) शून्य
(Zero) असते.

4) पूरक संच (Complementary Events): पूरक घटना अशा दोन घटना आहेत की एक घटना घडेल जर दुसरी घडली नाही. दोन घटनांना पूरक घटना म्हणून वर्गीकृत करण्यासाठी ते परस्पर अपवर्जी आणि सर्वसमावेशक असले पाहिजेत. उदाहरणार्थ, परीक्षा उत्तीर्ण होणे किंवा परीक्षा उत्तीर्ण न होणे. एखाद्या घटना (Events) वर्णन प्रयोगाच्या परिणामांचा संच म्हणून केले जाऊ शकते. अशा प्रकारे, घटना (Events) नेहमी नमुना अवकाश (Sample space) जागेचा उपसंच असेल. एक कार्यक्रम होऊ द्या. A ची पूरकता A' किंवा A^c म्हणून दर्शविली जाते. A आणि A' मिळून पूरक घटना तयार करतात.

5) परस्पर अपवर्जी घटना (Mutually exclusive Events): दोन किंवा अधिक संच परस्पर अपवर्जी आहेत असे म्हटले जाते, जर त्यांच्यात कोणताही साधा घटक सामाईक नसेल. दुसऱ्या शब्दांत सांगायचे तर, कोणत्याही घटनेची घटना इतर घटना वगळते. उदा. एक फासा टाकल्यानंतर
नमुना अवकाश (Sample space), S = {1, 2, 3, 4, 5, 6}
A आणि B या दोन घटना अशा आहेत
A={1, 3, 5} , B= { 2, 4, 6}
तर असे लक्षात येते की,
B= A'= संच A चा पूरक संच (Complement of set A)
A= B'= संच B चा पूरक संच (Complement of set B)
अशा प्रकारे A आणि B एकमेकांच्या पूरक घटना आहेत. A आणि B या दोन घटना परस्पर अपवर्जी आहेत असे म्हटले जाते जर A∩B=Ø, empty set.

6) सर्वसमावेशी घटना (Exhaustive events): दोन घटनांना परस्पर सर्वसमावेशी (Exhaustive) म्हटले जाते जर त्यापैकी किमान एक घटना घडण्याची खात्री असेल तर ती घटना निश्चितपणे घडेल.
उदा. छापा (Head) मिळण्याच्या घटना आणि काटा (Tail) मिळविण्याच्या घटना, जेव्हा नाणे फेकले जाते तेव्हा एकमेकांना सर्वसमावेशी असतात.

7) समसंभाव्य घटना)Equally likely Events): घटनांच्या समान संभाव्यतेसह दोन किंवा अधिक घटनांना समान संभाव्य घटना (Equally likely Events)म्हणतात. ते साधे किंवा मिश्रित असू शकत नाहीत
उदाहरण . एक नाणे फेकणे
A= छापा (H) मिळवण्याची घटना
B= काटा (T) मिळवण्याची घटना
या प्रयोगात, A आणि B दोन्ही घटना घडण्याची समान शक्यता आहे. जेणेकरुन A आणि B सारख्याच संभाव्य घटना आहेत. तितक्याच समसंभाव्य घटना (Equally likely Events)आणि परस्पर अपवर्जी घटना (Mutually exclusive Events) एकमेकांपासून स्वतंत्र आहेत.

सोडवलेली उदाहरणे (Solved Problems)

उदाहरण-3.1:

1. एक फासा **(dice)** फेकले असता. नमुना अवकाश **(Sample space)** लिहा
तसेच खालील घटना संच **(Events)** लिहा

a) फासा वरील संख्या सम **(Even no)** आहे

b) फासा वरील संख्या विषम **(Odd no)** आहे

या घटना समसंभाव्य घटना **(Equally likely Events)**आणि परस्पर अपवर्जी घटना **(Mutually exclusive Events)** आहेत का**?**

उत्तर : एक फासा फेकला असता

S = { 1, 2, 3, 4, 5, 6 }

घटना A = फासा वरील संख्या सम आहे

= { 2, 4, 6 }

घटना B = फासा वरील संख्या विषम आहे

= { 1, 3, 5 }

निरीक्षण केले असता असे आढळते की , A∩B=Ø, empty set.

घटना A आणि B परस्पर अपवर्जी घटना (Mutually exclusive Events) आहेत. ते एकाच वेळी घडू शकत नाही. तसेच A आणि B मध्ये प्रत्येकी 3 घटक आहेत. म्हणजेच ते समसंभाव्य घटना (Equally likely Events)आहेत कारण त्याची संभाव्यता सारखी आहे.

2) दोन नाणे फेकले असता नमुना अवकाश **(Sample space)** लिहा

तसेच खालील घटना संच **(Events)** लिहा

फक्त दोन छापा **(exactly 2 head)**

कमीत कमी एक छापा **(Atleast 1 head)**

एकही छापा नाही (No head)

उत्तर : दोन नाणे हवेत फेकले असता

नमुना अवकाश , S= { HH , HT, TH , TT }

a) घटना A = फक्त दोन छापा (exactly 2 head)

= { HH }

b) घटना B = कमीत कमी एक छापा (Atleast 1 head)

= { HT, TH, HH }

c) घटना C = एकही छापा नाही (No head)

= { TT }

संभाव्यतेची शास्त्रीय किंवा गणितीय व्याख्या (Definition of Probability)

एक सोपा प्रयोग विचारात घेऊ. एका पिशवीत समान आकाराचे चार चेंडू आ्हेत. त्यातील तीन चेंडू पांढरे व चौथा चेंडू काळा आ्हे. डोळे मिटून त्यातील एक चेंडू काढायचा आ्हे. काढलेला चेंडू पांढरा असण्याची शक्यता जास्त आ्हे, हे सहज कळते. **गणिती भाषेत एखाद्या अपेक्षित घटनेची शक्यता दर्शवणाऱ्या संख्येला संभव्यता (Probability) असे म्हणतात**.

एखाद्या याद्दच्छिक प्रयोगा साठी नमुना अवकाश S असेल आणि A ही त्या प्रयोगासंबंधी अपेक्षित घटना असेल तर त्या घटनेची संभाव्यता 'P(A)' अशी दर्शवतात आणि पुढील सूत्राने ठरवतात.

$$P(A) = \frac{n(A)}{n(S)} = \frac{\text{घटना } 'A' \text{ मधील नमुना घटकांची संख्या}}{\text{नमुना अवकाशातील } 'S' \text{ एकूण घटकांची संख्या}}$$

वरील प्रयोगात, उचललेला चेंडू पांढरा असणे' ही घटना A असेल, तर n(A) = 3, कारण पांढरे चेंडू तीन आहेत आणि एकूण चेंडू चार असल्याने $n(S) = 4$

उचललेला चेंडू पांढरा असणे, याची संभाव्यता $\mathrm{P(A)} = \frac{\mathrm{n(A)}}{\mathrm{n(S)}} = \frac{3}{4}$

तसेच 'उचललेला चेंडू काळा असणे' ही घटना B असेल, तर $n(B) = 1$

$$P(A) = \frac{n(A)}{n(S)} = \frac{1}{4}$$

सरावासाठी उदाहरणे (Excersice)

1. दोन नाणी फेकली असता खालील घटनांची संभाव्यता काढा.
   a. कमीत कमी एक छापा (H) मिळणे
   b. एकही छापा (H) न मिळणे.
2. दोन फासे एकाच वेळी टाकले असता खालील घटनांची संभाव्यता काढा.
   a. पृष्ठभागावरील अंकांची बेरीज कमीत कमी 10 असणे.
   b. पृष्ठभागावरील अंकांची बेरीज 33 असणे.
   c. पाहील्या फाशावरील अंक दुसऱ्या फाशावरील अंकापेक्षा मोठा असणे.
3. एका पेटीत 15 तिकटे आहेत. प्रत्येक टिकीटावर 1 ते 15 पैकी एक संख्या लिहीलेली आहे. त्या पेटीतून एक तिकीट या पद्धतीने काढले तर तिकीटावरची संख्या ही
   a. सम संख्या असणे.
   b. संख्या 5 च्या पटीत असणे, या घटनेची संभाव्यता काढा

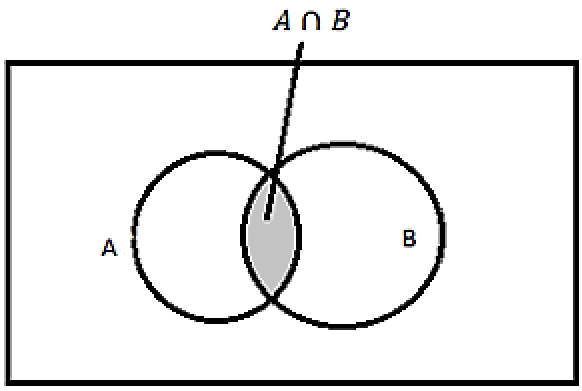
3.2 संभाव्यतेची प्रमेये- थिओरॅम्स ऑफ प्रोबॅबिलिटी (Theorems of Probability)

3.2.1 संभाव्यतेचे बेरजेचे सिद्धांत

व्याख्या: जर A आणि B या मर्यादित नमुना अवकाश S मधील दोन घटना असतील, तर घटनांपैकी किमान एक घडण्याची संभाव्यता अशी परिभाषित केली जाते,

$$P(A \cup B) \text{ or } P(A + B) = P(A) + P(B) - P(A \cap B)$$

Proof : घटना अवकाश (Sample space S) हा परस्पर अपवर्जी घटना (Mutually exclusive Events), सर्वसमावेशी घटना (Exhaustive events) , समसंभाव्य घटना (Equally likely Events) यांचा संचआ



वरील वेन आकृतीवरून (Venn diagram) , संच सिद्धांताची संकल्पना वापरून,

$n(A \cup B) = n(A) + n(B) - n(A \cap B)$

म्हणून $n(S) \neq 0$, दोनी बाजूला $n(S)$ ने Divide करा

$$\frac{n(A \cup B)}{n(S)} = \frac{n(A)}{n(S)} + \frac{n(B)}{n(S)} - \frac{n(A \cap B)}{n(S)}$$

संभाव्यतेच्या व्याख्ये नुसार,

$P(A \cup B) \text{ or } P(A + B) = P(A) + P(B) - P(A \cap B)$ -------------------- (1)

महत्वाचे टिप्स : जर A आणि B परस्पर अपवर्जी घटना असतील तर ते एकाच वेळी होऊ शकत नाहीत म्हणजेच ते रिक्त संच (Empty sets or Disjoint sets) असतात .

$A \cap B = 0$

$\therefore\ P(A \cap B) = 0$

म्हणून विधान (1) कमी होऊन बनले,

$P(A\ \cup B)\ or\ P(A + B\ ) = P(A) + P(B)$ --------------------- (2)

जर A , B आणि C परस्पर अपवर्जी घटना (mutually exclusive) आणि समसंभाव्य (Equally likely events)असतील तर

$$P(A\ \cup B \cup C)\ = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C) \quad \text{----------------------(3)}$$

जर A , B आणि C परस्पर अपवर्जी घटना (mutually exclusive events) असतील तर ,

$A \cap B = B \cap C = A \cap C = A \cap B \cap C = 0$

$\therefore P(A \cap B) = P(B \cap C) = P(A \cap C) = P(A \cap B \cap C) = 0$

म्हणून विधान (3) कमी होऊन बनले,

$P(A\ \cup B \cup C)\ = P(A) + P(B) + P(C)$ --------------------- (4)

3.2.2 $\emptyset$ हा जर रिक्त संच (Empty set) असेल तर $P(\emptyset) = 0$

जर $\emptyset$ हा जर रिक्त संच (Empty set)असेल तर

$\emptyset$ = empty set

$= \{\ \ \}$ $\therefore n(\emptyset) = 0$

संभाव्यतेच्या व्याख्ये नुसार,

$P(\emptyset) = \frac{n(\emptyset)}{n(S)}$ इथे $n(S) \neq 0$

$= \frac{0}{n(S)} = 0$

$\therefore\ \boldsymbol{P(\emptyset) = 0}$

3.2.3 कोणत्याही शून्य नसलेल्या घटना A (A$\neq$ 0) साठी, दाखवा $P(A') = 1 - P(A)$

संच A' हा संच A चा पूरक संच (Complementary set) आहे

संच सिद्धांतानुसार, $n(A' \cup A) = n(S)\ \&\ n(A' \cap A) = 0$

$P(A' \cup A) = \frac{n(A' \cup A)}{n(S)} = 1 \quad \& \quad P(A' \cap A) = 0$

संभाव्यतेच्या बेरजेच्या सिद्धांतानुसार,

$P(A\ \cup\ A')\ = P(A) + P(A') - P(A \cap A')$

$\therefore \quad 1 \quad = \quad P(A) + P(A') - 0$

$\therefore\ \ P(A) + P(A') = 1$ --------------------------------(5)

किंवा $P(A') = 1 - P(A)$ --- ---------- ------------------ (6)

संभाव्यतेच्या समस्या सोडवण्यासाठी उपयुक्त सूचना (Useful Hints for solving examples of probability) :

डी मॉर्गनचा (De Morgan's law) नियम संभाव्यतेचे उदाहरणे सोडवण्यासाठी खूप उपयोगी आहे.

$$(A \cup B)' =\ A'\ \cap\ B'$$
$$(A \cap B)' =\ A'\ \cup\ B'$$

कोणत्याही A आणि B दोन घटकासाठी

जर परस्पर अपवर्जी असतील तर $P(A\ \cup B)\ o = P(A) + P(B)$

जर सर्वसामावेशी असतील तर $P(A\ \cup B) = 1$

जर परस्पर अपवर्जी आणि सर्वसामावेशी असतील तर $P(A\ \cup B)\ = P(A) + P(B)$

कोणत्याही A आणि B दोन घटकासाठी

$$P(A) = P(A \cap B) + P(A \cap B')$$
$$P(B) = P(A \cap B) + P(A' \cap B)$$
$$P(A \cup B)' =\ P(A'\ \cap\ B')$$

$$P(A \cap B)' = P(A' \cup B')$$

संभाव्यतेमध्ये कमीत कमीचा अर्थ (Meaning of Atleast)

कमीत कमी एक घटक A आणि B मधील , म्हणजे $A \cup B$

$A \cup B = 1 - (A \cup B)'$

$\therefore P(A \cup B) = 1 - P(A \cup B)'$

$= 1 - P(A' \cap B')$

संभाव्यतेमध्ये जास्तीत जास्त चा अर्थ (Meaning of At most)

जास्तीत जास्त एक घटक A आणि B मधील, $A \cap B$

$A \cap B$ चा पूरक संच $(A \cap B)'$

$(A \cap B)' = A' \cup B'$

$\therefore P\ (A \cap B)' = P(A' \cup B')$

संभाव्यतेमध्ये ‘ फक्त ‘ चा अर्थ (meaning of only):

P(A आणि B पैकी फक्त एक) $= P(A \cap B') + P(A' \cap B)$

संयोजन/ कॉम्बिनेशन्स (Combinations)आणि क्रमपरिवर्तन/ पेरमुटेशन्स (Permutations) :

संयोजन/ कॉम्बिनेशन्स (Combination) : ${}^{n}C_{r} = \frac{n!}{r!(n-r)!}$

उदा. I) : ${}^{5}C_{1} = \frac{5!}{1!(5-1)!} = 5$

II) ${}^{10}C_{3} = \frac{10!}{3!(10-3)!} = 120$

Factorial n! $= \boldsymbol{n\,(n-1)(n-2)(n-3)\ldots\ldots 3 \times 2 \times 1}$

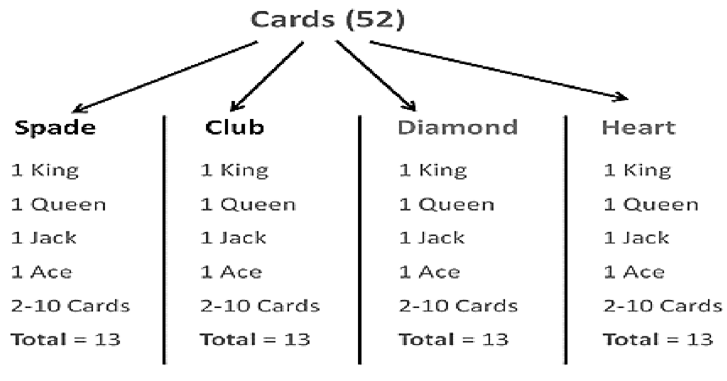
उदा. I) 5! = 5 × 4 × 3 × 2 × 1 = 120

II) $3\,! = 3 \times 2 \times 1 = 6$

क्रमपरिवर्तन/ पेरमुटेशन्स (Permutations) : ${}^{n}P_{r} = \frac{n!}{(n-r)!}$

उदा. I) : ${}^{5}P_{1} = \frac{5!}{(5-1)!} = 5$

52 पत्तेची पाने (52 playing cards) बदल माहिती:



एकूण पत्त्यांची पाने = 52 playing cards

एकूण नंबर असणारे कार्ड = 4$\times 10 = 40$ कार्ड

एकूण चित्र (Face cards) असणारे कार्ड = 4$\times 3 = 12$ कार्ड

एकूण राजा (King) = $4 \times 1 = 4$ कार्ड

एकूण राणी (Queen) = $4 \times 1 = 4$ कार्ड

एकूण जोकर (Joker) = $4 \times 1 = 4$ कार्ड

संभाव्यते मध्ये किंवा (OR) म्हणजे बेरीज (Adding the probability)

$P(A\ or\ B) = P(A) + P(B)$

संभाव्यते मध्ये आणि (AND) म्हणजे गुणाकार (Multiplying the probability)

$P(A\ and\ B) = P(A) \times P(B)$

सोडवलेली उदाहरणे (Solved Problems)

उदाहरण 3.2: तीन नाणी (coin)एकाच वेळी फेकली जातात. खालील घटकांची संभाव्यता (probability) काढा

a. सर्व छापा (H)
b. अगदी 1 छापा (H)
c. कमान 1 छापा (H)

उत्तर : घटना अवकाश S={HHH , HHT, HTH, TTH, THT, HTT, THH ,TTT } $n(S) = 8$

A= सर्व छापा (All Head)

= { HHH } $n(A) = 1$

$$\text{P(A)=P(All head)} = \frac{n(A)}{n(S)} = \frac{1}{8}$$

B= अगदी 1 छापा (exactly 1 head)

= { HTT, THT, TTH } $n(B) = 3$

$$\text{P(B)=P(exactly 1 head)} = \frac{n(B)}{n(S)} = \frac{3}{8}$$

C= कमान 1 छापा (atleast 1 head)

= 1 छापा किंवा 2 छापा किंवा 3 छापा

= { HHH, HHT, HTH, THH, HTT, THT, TTH} $n(C) = 7$

$$\text{P(C)=P(Atleast 1 head)} = \frac{n(C)}{n(S)} = \frac{7}{8}$$

उदाहरण 3.3: एक फासा टाकला असता खालील प्रत्येक अट पूर्ण करणाऱ्या घटनेची संभाव्यता काढा

a. वरच्या पृष्टभागावर 5
b. विषम संख्या (odd no)
c. संख्या 1 पेक्षा मोठी असणे (greater than 1)
d. संख्या 4 च्या पटीत असणे (multiple of 4)

उत्तर : घटना अवकाश $S = \{1, 2, 3, 4, 5, 6\}$ $n(S) = 6$

A= वरच्या पृष्टभागावर 5

= { 5 } $n(A) = 1$

$$\text{P(A)=P(5)} = \frac{n(A)}{n(S)} = \frac{1}{6}$$

B= विषम संख्या

= { 1, 3, 5 } $n(B) = 3$

$$\text{P(B)=P(odd no)} = \frac{n(B)}{n(S)} = \frac{3}{6} = \frac{1}{2}$$

C= संख्या 1 पेक्षा मोठी असणे

= { 2, 3, 4, 5, 6} $n(C) = 5$

$$\text{P(C)=P(greater than 1)} = \frac{n(C)}{n(S)} = \frac{5}{6}$$

D = संख्या 4 च्या पटीत असणे

= { 4 } $n(D) = 1$

$$\text{P(C)=P(multiple of 4)} = \frac{n(C)}{n(S)} = \frac{1}{6}$$

उदाहरण 3.4: कार्डांच्या पॅकमधून एक कार्ड काढले जाते. काढलेले कार्ड एकतर चौकट (diamond) किंवा किल्वर(Club) एक्का आहे याची संभाव्यता शोधा.

उत्तर :

A: चौकट कार्ड काढण्याची घटना

B : क्लबचा एक्का काढण्याची घटना.

चौकट (diamond) : P(A) चे कार्ड काढण्याची संभाव्यता : P(A) = $\frac{13}{52}$
किल्वर(Club) चा एक्का काढण्याची संभाव्यता. P(B) = $\frac{1}{52}$
परस्पर अपवर्जी घटना असल्यामुळे, चौकट (diamond) किंवा किल्वर(Club)
P(B) एक्का म्हणून कार्ड काढण्याची शक्यता आहे.

$$P(A \cup B) = P(A) + P(B) = \frac{13}{52} + \frac{1}{52} = \frac{14}{52}$$

उदाहरण 3.5: योग्य रीतीने पिसलेल्या ५२ पत्त्यांच्या कॅटमधून एक पत्ता काढला तर खालील घटनांची संभाव्यता काढा .
सात (7)
लाल पत्ता (a red colour card)
बदाम पत्ता (a heart)
लाल सहा (a red six)
उत्तर : योग्य रीतीने पिसलेल्या ५२ पत्त्यांच्या कॅटमधून एक पत्ता काढला तर $n(S) = 52$
A = सात (7)
= {4 cards) $n(A) = 4$

$$P(A) = \frac{n(A)}{n(S)} = \frac{4}{52} = \frac{1}{13}$$

B= बदाम पत्ता (a heart)
= {13 cards) $n(B) = 13$

$$P(B) = \frac{n(B)}{n(S)} = \frac{13}{52} = \frac{1}{4}$$

C= लाल पत्ता (a red colour card)
= {26 cards) $n(C) = 26$

P(C) $= \frac{n(C)}{n(S)} = \frac{26}{52} = \frac{1}{2}$

D= लाल सहा (a red six)
= {2 cards) $n(D) = 2$

$$P(D) = \frac{n(D)}{n(S)} = \frac{2}{52} = \frac{1}{26}$$

उदाहरण 3.6: एका वर्गात **20** मुले आणि **15** मुली आहेत. या विद्यार्थ्यांपैकी एकाची वर्गाचे प्रतिनिधित्व करण्यासाठी यादृच्छिकपणे निवड केली जाते. विद्यार्थी मुलगी निवडली जाण्याची संभाव्यता किती आहे **?**
उत्तर : एका वर्गात 20 मुले आणि 15 मुली आहेत. एकूण (20 मुले +15 मुली) = 35 विद्यार्थी
35 विद्यार्थ्यांपैकी एक यादृच्छिकपणे निवड केली जाते, ${}^{35}C_1$
A= विद्यार्थी मुलगी
= {${}^{15}C_1$ } $n(A) = 15$

$$P(A) = \frac{n(A)}{n(S)} = \frac{15}{35} = \frac{3}{7}$$

उदाहरण 3.7: एका बॉक्समध्ये **8** निळे**, 7** हिरवे आणि **5** लाल बॉल असतात. बॉक्समधून एक बॉल घेतला जातो. बॉल असण्याची संभाव्यता काय आहे**?**

a. लाल बॉल
b. निळा बॉल
c. हिरवे बॉल
d. पिवळा बॉल

उत्तर : एका बॉक्समध्ये एकूण (8 निळे + 7 हिरवे +5 लाल बॉल)= 20 बॉल बॉक्स मधून एक बॉल निवडला तर तो २० पद्धतीने निवडता येईल. $n(s) = 20$

a) A = लाल बॉल

= { 5 लाल बॉल}

$n\ (A) = 5$

$$P(A) = \frac{n(A)}{n(S)} = \frac{5}{20} = \frac{1}{4}$$

b) B = निळा बॉल

= { 8 निळा बॉल } $n(B) = 8$

$$P(B) = \frac{n(B)}{n(S)} = \frac{8}{20} = \frac{2}{5}$$

c) A = हिरवे बॉल

= { 7 हिरवे बॉल ल} $n\ (C) = 7$

$$P(C) = \frac{n(C)}{n(S)} = \frac{7}{20}$$

d) D = पिवळा बॉल

= { 0 पिवळा बॉल } $n\ (D) = 0$

$$P(D) = \frac{n(D)}{n(S)} = \frac{0}{20} = 0$$

उदाहरण 3.8: एका बॅगमध्ये **1** ते **20** क्रमांकाची **20** तिकिटे असतात**.** एक तिकीट यादृच्छिकपणे काढले जाते**.** संभाव्यता शोधा की ती **3** किंवा **5** च्या पटीने क्रमांकित आहे**.**

उत्तर : एका बॅगमध्ये 1 ते 20 क्रमांकाची 20 तिकिटे आहेत. त्यापैकी एक तिकीट सिलेक्ट केले. $^{35}C_1$ पद्धतीने म्हणजेच २० ways.

घटना अवकाश $S = \{1, 2, 3, 4, \ldots\ldots, 19, 20\}$ $n(S) = 20$

A= तिकीट वरील नंबर हा 3 किंवा 5 च्या पटीत आहे(multiple of 3 or 5)

= { 3, 5, 6, 9, 10, 12, 15, 18, 20 } } $n(A) = 9$

$$P(A) = \frac{n(A)}{n(S)} = \frac{9}{20}$$

उदाहरण 3.9: योग्य रीतीने पिसलेल्या **52** पत्त्यांच्या कॅटमधून यादृच्छिकपणे काढलेली दोन कार्डे काढली असता एकाच सूटच्या राजा **(king)** आणि राणीच्या **(queen)** जोडी मिळण्याची शक्यता शोधा**.**

उत्तर : 52 पत्त्यांच्या कॅटमधून 2 कार्ड बाहेर $^{52}C_2$ ways ने काढता येते.

$$n(S) = {}^{52}C_2 = \frac{52\times51}{1\times2} = 26 \times 51$$

घटना A = एकाच सूटच्या राजा (king) आणि राणीच्या (queen) जोडीतून काढलेली दोन कार्डे

= प्रत्येक सूट मध्ये राजा (king)आणि राणी (queen)ची एक जोडी असते आणि असे 4 सूट (pair) असतात.

= { 4 pair }

$n(A) = 4$

$$P(A) = \frac{n(A)}{n(S)} = \frac{4}{26 \times 51} = \frac{2}{663}$$

उदाहरण 3.10: मुकेश आणि राकेश या दोन विद्यार्थ्यांद्वारे उदाहरणे सोडवण्याची संभाव्यता $\frac{1}{2}$ आणि $\frac{1}{3}$ आहे**.** समस्येचे निराकरण होण्याची शक्यता काय आहे**?**

उत्तर : A = मुकेशने समस्या सोडवणे

$P(A) = \frac{1}{2}$

B = राकेशने समस्या सोडवणे

$P(B) = \frac{1}{3}$

समस्येचे निराकरण होण्याची संभाव्यता म्हणजेच
समस्या सोडवली जाते, जर ती त्यापैकी किमान एकाने सोडवली असेल
समजा E= घटना " समस्येचे निराकरण होणे"
= A किंवा B किंवा दोघे

$P(E) = P\left(\text{ A किंवा B किंवा दोघे}\right)$

$= P(A \cup B)$

$= P(A) + P(B) - P(A \cap B)$

$= P(A) + P(B) - P(\text{ A आणि B})$

$= P(A) + P(B) - P(A) \times (B)$

$= \frac{1}{2} + \frac{1}{3} - (\frac{1}{2} \times \frac{1}{3})$

$= \frac{4}{6}$

$= \frac{2}{3}$

सरावसंच:

1) एकाच वेळी तीन नाणी फेकली असता खालील घटनांची संभाव्यता शोधा
   a) 3 छापा (3 head)
   b) किमान 1 छापा (atleast 1 head)
   c) किमान 2 छापा (atleast 2 head)
2) दोन नाणी फेकली असता खालील घटनांची संभाव्यता काढा.
   a) कमीत कमी एक छापा (H) मिळणे
   b) एकही छापा (H) न मिळणे.
3) 2 फासे एकाच वेळी फेकले असता पृष्ठभागावर दिसणाऱ्या खालील घटनांची संभाव्यता काढा.
   a) एकूण बेरीज 6
   b) एकूण बेरीज 9 पेक्षा जास्त
   c) एकूण बेरीज 7 पेक्षा कमी
4) एका टेनिस क्लब मध्ये 16 मुली (girls) आणि 8 मुले (boys) आहेत. यापैकी एक याद्दचिक पद्धतीने निवडला असता ती मुलगी असेल याची संभाव्यता किती असेल?
5) एका बॉक्समध्ये 7 निळे (blue) , 8 हिरवे (green) , 5 नारंगी (orange) आणि 10 लाल (red) बॉल असतात. बॉक्समधून एक बॉल घेतला असता तो बॉल खालील पैकी असण्याची संभाव्यता काय आहे?
   a) लाल बॉल (red ball)
   b) नारंगी बॉल (orange ball)
   c) हिरवे किंवा लाल बॉल (green or red ball)
   d) निळा बॉल नाही (not blue ball)
6) 2 फासे एकाच वेळी फेकले असता पृष्ठभागावर दिसणाऱ्या संख्येची एकूण बेरीज 9 आहे त्याची संभाव्यता काढा.
7) एक फासा टाकला असता वरच्या पृष्टभागावरील संख्या 4 पेक्षा मोठी असण्याची संभाव्यता काढा.
8) योग्य रीतीने पिसलेल्या ५२ पत्त्यांच्या कॅटमधून एक पत्ता काढला तर असता कलर कार्ड (Colour card) असण्याची संभाव्यता काढा.
9) एका वर्गखोलीमध्ये 12 विद्यार्थी आहेत त्या पैकी 5 मुले आणि बाकीचे मुली आहेत तर यापैकी एक याद्दचिक पद्धतीने निवडला असता ती मुलगी असेल याची संभाव्यता किती असेल?

10) एका बॉक्समध्ये 5 पांढरे (white) , 8 हिरवे (green) आणि 7 लाल (red) बॉल असतात. बॉक्समधून एक बॉल घेतला असता तो बॉल पांढरा नसण्याची संभाव्यता काय आहे?

11) एका बॅग मध्ये 10 पांढरे आणि 10 काळे बॉल्स (Balls) आहेत. त्यापैकी 2 बॉल्स (Balls) बाहेर काढले असता ते एकाच रंगाचे (same colour) असण्याची संभाव्यता काढा .

12) जर A आणि B या मर्यादित नमुना अवकाश S मधील दोन घटना असतील, $P(A \cup B) = \frac{5}{6}$, $P(A \cap B) = \frac{1}{3}$, $P(B') = \frac{1}{3}$, तर $P(A)$ किंमत काढा

13) जर $P(A) = \frac{1}{2}$, $P(A \cap B) = \frac{1}{6}$ & $P(A \cup B) = \frac{7}{12}$, तर $P(B)$ & $P(A' \cup B')$किंमत काढा.

14) योग्य रीतीने पिसलेल्या ५२ पत्त्यांच्या कॅटमधून एक पत्ता काढला तर काढलेला कार्ड एक्का (ace) किंवा राजा (king) असण्याची संभाव्यता किती असेल?

15) 2 फासे एकाच वेळी फेकले असता पृष्ठभागावर दिसणाऱ्या संख्येची एकूण बेरीज ७ किंवा गुणाकार १२ असणाऱ्या घटनेची संभाव्यता काढा.

उत्तरे

1) a) $\frac{1}{8}$ b) $\frac{7}{8}$ c) $\frac{1}{2}$

2) a) $\frac{3}{4}$ b) $\frac{1}{4}$

3) a) $\frac{5}{36}$ b) $\frac{1}{6}$ c) $\frac{5}{12}$

4) $\frac{2}{3}$

5) a) $\frac{1}{3}$ b) $\frac{1}{6}$ c) $\frac{3}{5}$ d) $\frac{23}{30}$

6) $\frac{1}{9}$ 7) $\frac{1}{3}$ 8) 1 9) $\frac{7}{12}$

10) $\frac{3}{4}$ 11) $\frac{9}{19}$ 12) $\frac{1}{2}$

13) a) $\frac{1}{4}$ b) $\frac{5}{6}$ 14) $\frac{2}{13}$ 15) $\frac{2}{9}$

3.3 सशर्त संभाव्यता/ कंडिशनल प्रोबॅबिलिटी (CONDITIONAL PROBABILITY): दैनंदिन व्यवहारात अगदी सहजपणे संभाव्यतेविषयी बोलले जाते. उदाहरणार्थ, आज पाऊस पडण्याची शक्यता किती आहे?, भारतीय संघ उद्याच्या क्रिकेटच्या सामन्यात जिंकण्याची शक्यता किती आहे? अशा स्वरूपाच्या चर्चा या संभाव्यतेविषयी असतात. 'सशर्त संभाव्यता' हा संभाव्यतेचाच एक प्रकार आहे. समजा आकाशात भरपूर काळे ढग दिसत असतील, तर पाऊस पडण्याची शक्यता जवळजवळ १००% आहे, असे म्हटले जाते. आकाशात काळे ढग दिसणे या घटनेमुळे पाऊस पडण्याच्या शक्यतेत वाढ होणे हे सशर्त संभाव्यतेचेच उदाहरण आहे. 'A' या घटनेची संभाव्यता काढताना, दुसरी 'B' ही घटना घडली असल्याची माहिती वापरून काढलेल्या संभाव्यतेस 'A' ची सशर्त संभाव्यता असे म्हणतात. $P(A \backslash B)$ या संकेतनाने ही सशर्त संभाव्यता दर्शविली जाते.सशर्त संभाव्यता काढण्यासाठी दोन घटनांच्या छेदनाच्या (एकत्रितपणे घडून येण्याच्या) संभाव्यतेचा वापर केला जातो. सशर्त संभाव्यता काढण्याचे सूत्र पुढील प्रमाणे आहे.

$$P(A \backslash B) = P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)} \ , \ P(B) \neq 0$$

वरील समीकरणात 'B' ही अशक्य घटना नाही, म्हणजेच $P(B)$ शून्य नाही असे मानले जाते.

$$P(B \backslash A) = P\left(\frac{B}{A}\right) = \frac{P(A \cap B)}{P(A)} \ , \ P(A) \neq 0$$

3.3.1 स्वतंत्र घटना (INDEPENDENT EVENTS): दोन किंवा अधिक घटना स्वतंत्र आहेत असे म्हटले जाते, जर त्यांपैकी कोणत्याही एकाची संभाव्यता इतर कोणत्याही घटनांच्या घटनेमुळे प्रभावित होत नसेल.

$$\boldsymbol{P}\left(\frac{A}{B}\right) = \boldsymbol{P}(\boldsymbol{A}) \quad \text{आणि} \quad \boldsymbol{P}\left(\frac{B}{A}\right) = \boldsymbol{P}(\boldsymbol{B})$$

∴ स्वातंत्र्य म्हणजे B चे निरीक्षण केल्याने A च्या संभाव्यतेवर कोणताही परिणाम होत नाही.

3.3.2 संभाव्यतेचे गुणाकाराचे सिद्धांत (MULTILICATION THEOREM OF PROBABILITY)

जर A आणि B दोन स्वतंत्र घटना असतील , तर $\boldsymbol{P}(\boldsymbol{A} \cap \boldsymbol{B}) = \boldsymbol{P}(\boldsymbol{A}) \times \boldsymbol{P}(\boldsymbol{B})$ ------------------- - (1)

याला संभाव्यतेचे गुणाकार प्रमेय (Multiplication theorem of Probability) असे म्हणतात.

विधान तीन स्वतंत्र घटना A, B, C साठी वाढवले जाऊ शकते.

$\boldsymbol{P}(\boldsymbol{A} \cap \boldsymbol{B} \cap \boldsymbol{C}) = \boldsymbol{P}(\boldsymbol{A}) \times \boldsymbol{P}(\boldsymbol{B}) \times \boldsymbol{P}(\boldsymbol{C})$ -------------------- (2)

<u>टिप :</u>

1) जर A आणि B दोन स्वतंत्र घटना असतील तर

a) A आणि B'

b) A' आणि B

c) A' आणि B' देखील स्वतंत्र घटना आहेत.

2) दोन घटना A आणि B स्वतंत्र आहेत जर $P(A \cap B) = P(A) \times P(B)$

अन्यथा ते एकमेकांवर अवलंबून आहेत.

सोडवलेली उदाहरणे (Solved Problems)

उदाहरण 3.11: जर $\boldsymbol{P}(\boldsymbol{A}) = \frac{2}{3}$, $\boldsymbol{P}(\boldsymbol{B}') = \frac{3}{4}$ आणि $\boldsymbol{P}\left(\frac{A}{B}\right) = \frac{4}{5}$, तर $\boldsymbol{P}(\boldsymbol{A} \cap \boldsymbol{B})$ आणि $\boldsymbol{P}\left(\frac{B}{A}\right)$ संभाव्यता काढा.

उत्तर: $P(A) = \frac{2}{3}$, $P(B') = \frac{3}{4}$

$\mathrm{P(B)} = 1 - \mathrm{P(B')} = 1 - \frac{3}{4} = \frac{1}{4}$

$\mathrm{P}\left(\frac{\mathrm{A}}{\mathrm{B}}\right) = \frac{4}{5}$

a) सशर्त संभाव्यतेनुसार, $P(A \cap B) = P(B) \cdot P\left(\frac{A}{B}\right) = \frac{1}{4} \times \frac{4}{5} = \frac{1}{5}$

b) $C = P(B) \cdot P\left(\frac{B}{A}\right)$

$P\left(\frac{B}{A}\right) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{5}}{\frac{2}{3}} = \frac{3}{10}$

उदाहरण 3.12: बॉक्स A मध्ये 2 सिल्वर (Silver) आणि 4 कॉपर (copper)नाणी आहेत. बॉक्स B मध्ये 4 सिल्वर(Silver) आणि 3 कॉपर (copper) नाणी आहेत. यादृच्छिकपणे एक बॉक्स निवडला जातो आणि यादृच्छिकपणे त्यातून नाणे काढले जाते. ते कॉपरचे (copper) नाणे असण्याची शक्यता किती आहे?

उत्तर: समाज A = 1^{st} बॉक्स निवडला आणि त्यातील कॉपर नाणे निवडले

$P(A) = \frac{1}{2} \times \frac{4}{6} = \frac{1}{3}$

B =2^{nd} बॉक्स निवडला आणि त्यातील कॉपर नाणे निवडले

$P(B) = \frac{1}{2} \times \frac{3}{7} = \frac{3}{14}$

इथे, A आणि B परस्पर अपवारजी घटना आहेत,

$\therefore P(A \cup B) = P(A) + P(B) = \frac{1}{3} + \frac{3}{14} = \frac{23}{42}$

उदाहरण 3.13: दोन विद्यार्थी स्पर्धा जिंकण्याचे शक्यता अनुक्रमे $\frac{1}{2}$ आणि $\frac{1}{3}$.जर त्यांनी एकाच कंडिशन मध्ये स्पर्धे मध्ये सहभाग घेतला तर त्यांची जिंकण्याची संभाव्यता काढा.

उत्तर: $P(A) = \frac{1}{2}$ $\qquad P(B) = \frac{1}{3}$

$P(A') = 1 - P(A)$ $\qquad P(B') = 1 - P(B)$

$= 1 - \frac{1}{2} = \frac{1}{2}$ $\qquad = 1 - \frac{1}{3} = \frac{2}{3}$

कोणतीही एक घटना जिंकण्याची शक्यता $A \cup B$

$\therefore \; P(A \cup B) = 1 - P(A \cup B)'$

$= P(A' \cap B')$

$= 1 - P(A') \cdot P(B')$

$= 1 - \left(\frac{1}{2} \times \frac{2}{3}\right)$

$= \frac{2}{3}$

उदाहरण 3.14: योग्य रीतीने पिसलेल्या 52 पत्त्यांच्या कॅटमधून याद्दच्छिकपणे काढलेली दोन कार्डे काढली असता पहिले कार्ड राजा (king) आणि दुसरे राणी (queen)असण्याची शक्यता शोधा.

जर a) पहिले कार्ड बदलले असेल

b) बदलले नाही.

उत्तर: a) पहिले कार्ड बदलले असेल,

52 पत्त्यांच्या कॅटमध्ये ४ राजा (king) असता

∴P(पहिले कार्ड राजा (king)) = $\frac{4}{52} = \frac{1}{13}$

कार्ड बदलले जाते आणि नंतर राणी (queen) कार्ड काढले जाते.

∴ P(दुसरे कार्ड राणी(queen)) = $\frac{4}{52} = \frac{1}{13}$

दोन घटना संभाव्यतेच्या सशर्त आणि गुणाकार प्रमेयचे अनुसरण करतात

∴ P(पहिले कार्ड राजा आणि दुसरे कार्ड राणी)

= P(पहिले कार्ड राजा) × P(दुसरे कार्ड राणी)

$= \frac{1}{13} \times \frac{1}{13}$

$= \frac{1}{169}$

b) पहिले कार्ड न बदलता

∴ P(पहिले कार्ड राजा (king)) = $\frac{4}{52} = \frac{1}{13}$

पहिले कार्ड न बदलता

∴ P(दुसरे कार्ड राणी(queen)) = $\frac{4}{51}$

∴ P**(** पहिले कार्ड राजा आणि दुसरे कार्ड राणी **)**

= **P(** पहिले कार्ड राजा **) × P(**दुसरे कार्ड राणी **)**

$= \frac{\mathbf{1}}{\mathbf{13}} \times \frac{\mathbf{4}}{\mathbf{51}}$

$= \frac{4}{663}$

उदाहरण 3.15: A, B आणि C या तीन विद्यार्थ्यांना गणिताची समस्या दिली जाते ज्यांच्या सोडवण्याची शक्यता अनुक्रमे $\frac{1}{2}$,$\frac{3}{4}$ आणि $\frac{1}{4}$ आहे. संभाव्यता काय आहे

a) समस्या सुटणार का?

b) प्रत्येकाने समस्या सोडवली आहे का?

c) समस्या त्यांच्यापैकी एकाने सोडवली नाही का?

उत्तर: $P(A) = \frac{1}{2}$ $\therefore P(A') = 1 - P(A) = 1 - \frac{1}{2} = \frac{1}{2}$

$P(B) = \frac{3}{4}$ $\therefore P(B') = 1 - P(B) = 1 - \frac{3}{4} = \frac{1}{4}$

$P(C) = \frac{1}{4}$ $\therefore P(C') = 1 - P(C) = 1 - \frac{1}{4} = \frac{3}{4}$

a) समस्या सोडवली जाईल, जर त्यापैकी किमान एकाने समस्या सोडवली.
याचा अर्थ आपल्याला शोधावे लागेल , $P(A \cup B \cup C)$

$P(A \cup B \cup C) = 1 - P(A \cup B \cup C)'$

$= 1 - P(A' \cap B' \cap C')$

$= 1 - P(A') \times P(B') \times P(C')$

$= 1 - \frac{1}{2} \times \frac{1}{4} \times \frac{3}{4}$

$= 1 - \frac{3}{32}$

$= \frac{29}{32}$

b) प्रत्येकाने समस्या सोडवली ,

P (प्रत्येकाने समस्या सोडवली) = P(A ∩ B ∩ C)

जर A, B & C हे स्वतंत्र घटना (independent events) असतील , तर

$P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$

$= \frac{1}{2} \times \frac{1}{4} \times \frac{3}{4}$

$= \frac{3}{32}$

c) समस्या त्यांच्यापैकी एकाने सोडवली नाही

P (समस्या त्यांच्यापैकी एकाने सोडवली नाही) $= P(A' \cap B' \cap C')$

जर A, B & C हे स्वतंत्र घटना (independent events) असतील , तर A', B' & C' हे सुद्धा स्वतंत्र घटना (independent events) असतील.

$P(A' \cap B' \cap C') = P(A') \times P(B') \times P(C')$

$= \frac{1}{2} \times \frac{1}{4} \times \frac{3}{4}$

$= \frac{3}{32}$

सरावसंच (Excersice)

1) जर A आणि B या नमुना स्पेस S मधील दोन घटना आहेत जसे की $P(A) = 0.8$ आणि $P(B) = 0.6$ आणि $P(A \cup B) = 0.9$ तर खालील घटनांची संभाव्यता काढा.

(a) $P(A \cap B)$, $(b)\ P\left(\frac{A}{B}\right)$ $(c)\ P\left(\frac{B}{A}\right)$

2) जर $P(A) = \frac{1}{2}$, $P(B') = P(A \cup B) = \frac{2}{3}$ खालील घटनांची संभाव्यता काढा

(a) $P(A' \cap B')$ $(b)\ P\left(\frac{A}{B}\right)$

3) एक फासा (die) टाकल्या नंतर

घटना A = पृष्ठभागावरील नंबर विषम (Odd) आहे

घटना B = पृष्ठभागावरील नंबर ३ ने लहान (less than 3) आहे
तर A आणि B स्वतंत्र आहेत हे दाखवा.

4) बॉक्स A मध्ये 4 सिल्वर (Silver) आणि 5 कॉपर (copper)नाणी आहेत. बॉक्स B मध्ये 3 सिल्वर(Silver) आणि 4 कॉपर (copper) नाणी आहेत. यादृच्छिकपणे एक बॉक्स निवडला जातो आणि यादृच्छिकपणे त्यातून नाणे काढले जाते. ते कॉपरचे (copper) नाणे असण्याची शक्यता किती आहे?

5) A, B, C ही समस्या सोडवण्याची संभाव्यता अनुक्रमे $\frac{1}{3}, \frac{4}{5}, \frac{2}{7}$ आहे. त्या सर्वांनी स्वतंत्रपणे प्रयत्न केले तर. समस्या सोडवली जाण्याची शक्यता शोधा.

6) एका बॉक्समध्ये 3 लाल आणि 2 पांढरे बॉल असतात. दुसऱ्या बॉक्समध्ये 2 लाल आणि 4 पांढरे बॉल आहेत. पहिल्या बॉक्समधून एक बॉल यादृच्छिकपणे निवडला जातो आणि दुसऱ्या बॉक्समध्ये पाठविला जातो, त्यानंतर दुसऱ्या बॉक्समधून यादृच्छिकपणे एक चेंडू काढला जातो. तो लाल बॉल असल्याची संभाव्यता शोधा.

7) एक नाणे आणि फासा एकाच वेळी फेकले असता घटना छापा (Head) आणि 6 स्वतंत्र (independent) आणि परस्पर अपवर्जी (mutually exclusive) आहेत हे दाखवा.

<u>उत्तरे</u>

1) (a) $\frac{1}{2}$, (b) $\frac{5}{6}$, (c) $\frac{5}{8}$ 2) (a) $\frac{1}{3}$, (b) $\frac{1}{2}$

4) $\frac{71}{126}$ 5) $\frac{19}{21}$ 6) $\frac{13}{25}$

3.4 बायेस सिद्धांत (Bayes' Theorem)

प्रमेय विधान: जर A आणि B हे दोन घटना असतील तर बेजचे च्या सिद्धांतानुसार,

$$\boldsymbol{P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}} \ , \ \boldsymbol{P(B) \neq 0}$$

जर A आणि B हे दोन घटना असतील आणि $P(B) \neq 0$

* $P(A|B)$हि सशर्त संभाव्यता आहे : - A या घटनेची संभाव्यता काढताना, दुसरी 'B' ही घटना घडली असल्याची माहिती वापरली जाते

* $P(B|A)$हि सशर्त संभाव्यता आहे : - B या घटनेची संभाव्यता काढताना, दुसरी 'A' ही घटना घडली असल्याची माहिती वापरली जाते.

A ची शक्यता B निश्चित केली आहे किंवा दिली आहे म्हणून त्याचा अर्थ लावला जाऊ शकतो कारण

$$P(A|B) = P(B|A)$$

पुरावा:

जर A आणि B हे दोन घटना असतील तर

$$P(A|B) = P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)} \ , \ P(B) \neq 0 \qquad \text{-----} (1)$$

जेथे $P(A \cap B)$ ही A आणि B दोन्ही सत्य असण्याची संभाव्यता आहे.

त्याचप्रमाणे $P(B|A) = P\left(\frac{B}{A}\right) = \frac{P(A \cap B)}{P(A)} \ , \ P(A) \neq 0$ ------(2)

$$\therefore \ P(A \cap B) = P(B|A) \cdot P(A) \qquad \text{-----} (3)$$

समीकरण (3) च्या माहितीवरून , समीकरण (1) ची किंमत मिळेल,

$$\boldsymbol{P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}} \ , \ \boldsymbol{P(B) \neq 0}$$

सोडवलेली उदाहरणे (Solved Problems)

उदाहरण 3.15: तीन बॉक्समध्ये अनुक्रमे 6 लाल, 4 काळे ; 4 लाल, 6 काळे आणि 5 लाल, 5 काळे बॉल आहेत . बॉक्सपैकी एक याद्दच्छिकपणे निवडला जातो आणि त्यातून एक बॉल काढला जातो. काढलेला चेंडू लाल असल्यास, तो पहिल्या बॉक्समधून काढल्याची संभाव्यता शोधा.

उत्तर: X, Y, Z आणि A ची व्याख्या खालीलप्रमाणे करू द्या

X = प्रथम बॉक्स निवडला आहे

Y= बॉक्स दुसरा निवडला आहे

Z =बॉक्स तिसरा निवडला आहे

A = काढलेला बॉल लाल आहे

तीन बॉक्स असल्यामुळे आणि तीन बॉक्सपैकी एक याद्दच्छिकपणे निवडला आहे, म्हणून:

$$P(X) = P(Y) = P(Z) = 1/3$$

जर X आधीच आला असेल, तर प्रथम बॉक्स निवडला गेला आहे, ज्यामध्ये 6 लाल आणि 4 काळे बॉल आहेत. त्यातून लाल बॉल काढण्याची शक्यता $6/10$ आहे.

तर, $P(A/X) = 6/10$

त्याचप्रमाणे, तुमच्याकडे $P(A/Y) = 4/10$ आणि $P(A/Z) = 5/10$ आहेत

तुम्हाला P(X/A) शोधणे आवश्यक आहे, म्हणजे, काढलेला बॉल लाल आहे, तो पहिल्या बॉक्सतमधून काढला जाण्याची शक्यता किती आहे.

बायेस प्रमेयानुसार,

$$P\left(\frac{X}{A}\right) = \frac{P(X)\,P(A/X)}{P(X)\,P(A/X) + P(Y)\,P(A/Y) + P(Z)\,P(A/Z)}$$

$$= \frac{\frac{1}{3}\times\frac{6}{10}}{\left(\frac{1}{3}\times\frac{6}{10}\right) + \left(\frac{1}{3}\times\frac{4}{10}\right) + \left(\frac{1}{3}\times\frac{5}{10}\right)}$$

$$= \frac{2}{5}$$

उदाहरण 3.16: एका विमा कंपनीने 2000 स्कूटर चालक, 4000 कार चालक आणि 6000 ट्रक चालकांचा विमा उतरवला. स्कूटर चालक, कार चालक आणि ट्रक यांचा समावेश असलेल्या अपघाताची संभाव्यता अनुक्रमे 0.01, 0.03 आणि 0.015 आहे. विमाधारक व्यक्तींपैकी एकाला अपघात होतो. तो स्कूटर चालक असण्याची शक्यता किती आहे?

उत्तर: X ,Y, Z आणि A ची व्याख्या खालीलप्रमाणे करू द्या:

X = निवडलेली व्यक्ती स्कूटर चालक आहे

Y = निवडलेली व्यक्ती कार चालक आहे

Z = निवडलेली व्यक्ती ट्रक चालक आहे आणि

A = व्यक्तीला अपघात झाला

12000 लोक असल्यामुळे:

$P(X) = 2000/12000 = 1/6$

$P(Y) = 4000/12000 = 1/3$

$P(Z) = 6000/12000 = 1/2$

असे दिले आहे की,

P(A/X) = एखादी व्यक्ती स्कूटर चालक आहे हे लक्षात घेऊन अपघात होण्याची शक्यता = 0.01

त्याचप्रमाणे, तुमच्याकडे $P(A/Y) = 0.03$ आणि $P(A/Z) = 0.15$ आहे.

तुम्हाला $P(X/A)$ शोधणे आवश्यक आहे, म्हणजे त्या व्यक्तीला अपघात झाला आहे, तो स्कूटर चालक असण्याची शक्यता किती आहे?

$$P(X/A) = \frac{P(X)\,P(A/X)}{P(X)\,P(A/X) + P(Y)\,P(A/Y) + P(Z)\,P(A/Z)}$$

$$= \frac{\frac{1}{6} \times 0.01}{(1/6 * 0.01) + (1/3 \times 0.03) + (1/2 \times 0.15)}$$

$$= \frac{1}{52}$$

उदाहरण 3.17: एका विशिष्ट परिसरात, 90% मुले फ्लूमुळे आणि 10% गोवरमुळे आजारी पडली, इतर कोणत्याही आजाराची नोंद नाही. गोवरसाठी पुरळ दिसण्याची शक्यता ०.९५ आणि फ्लूसाठी ०.०८ आहे. जर एखाद्या मुलास पुरळ उठत असेल तर, मुलाला फ्लू होण्याची शक्यता शोधा.

उत्तर:

चला खालील घटना दर्शवूया:

F: फ्लू असलेली मुले

M: गोवर असलेली मुले

R: पुरळ झाल्याचे लक्षण दर्शवणारी मुले

आम्ही खालील संभाव्यता मोजू शकतो:

$$P(F) = 90\% = 0.9$$
$$P(M) = 10\% = 0.1$$
$$P(R|F) = 0.08$$
$$P(R|M) = 0.95$$

$$P(F|R) = \frac{P(R|F)P(F)}{P(R|M)P(M) + P(R|F)P(F)}$$

$$P(F|R) = \frac{0.08 \times 0.9}{0.95 \times 0.1 + 0.08 \times 0.9}$$

$$= 0.43$$

म्हणून, $P(F|R) = 0.43$

सरावसंच (Excersice)

a. A, B आणि C या तीन व्यक्तींनी एका खाजगी कंपनीत नोकरीसाठी अर्ज केला आहे. त्यांच्या निवडीची शक्यता $1:2:4$ च्या प्रमाणात आहे. A, B, आणि C कंपनीच्या नफ्यात वाढ करण्यासाठी बदल घडवून आणू शकतील अशा संभाव्यता अनुक्रमे $0.8, 0.5$ आणि 0.3 आहेत. इच्छित बदल होत नसल्यास, C च्या नियुक्तीमुळे होण्याची शक्यता शोधा.

b. एका पिशवीत 4 चेंडू असतात. दोन चेंडू बदलीशिवाय यादृच्छिकपणे काढले जातात आणि दोन्ही निळे असल्याचे आढळले. पिशवीतील सर्व चेंडू निळे असण्याची शक्यता किती आहे?

c. 52 कार्डांच्या डेकमधून एक कार्ड हरवले आहे. उर्वरित कार्डांमधून यादृच्छिकपणे दोन कार्डे काढली जातात आणि दोन्ही किल्व्हर असल्याचे आढळले. गमावले कार्ड देखील एक किल्व्हर आहे संभाव्यता काय आहे?

d. एका विमा कंपनीने 4000 डॉक्टर, 8000 शिक्षक आणि 12000 व्यावसायिकांचा विमा उतरवला आहे. डॉक्टर, शिक्षक आणि व्यावसायिक 58 वर्षांच्या आधी मरण पावण्याची शक्यता अनुक्रमे $0.01, 0.03$ आणि 0.05 आहे. जर विमाधारक व्यक्तींपैकी एकाचा 58 वर्षापूर्वी मृत्यू झाला तर तो डॉक्टर असल्याची शक्यता शोधा.

e. एका बॅग A मध्ये 4 पांढरे आणि 6 काळे बॉल असतात तर दुसऱ्या बॅग B मध्ये 4 पांढरे आणि 3 काळे बॉल असतात. एका पिशवीतून एक बॉल यादृच्छिकपणे काढला जातो आणि तो काळा असल्याचे आढळले आहे. ती बॅग A मधून काढली असल्याची संभाव्यता शोधा.

f. एक माणूस 3 पैकी 2 वेळा सत्य बोलण्यासाठी ओळखला जातो. तो फासा फेकतो आणि नोंदवतो की मिळालेली संख्या चार आहे. मिळालेली संख्या प्रत्यक्षात चार असल्याची संभाव्यता शोधा.

उत्तरे

a) $\frac{7}{10}$ b) $\frac{3}{5}$ c) $\frac{11}{50}$ d) $\frac{1}{22}$

e) $\frac{7}{12}$ f) $\frac{2}{7}$

संदर्भ (Reference):

1. https://archive.nptel.ac.in/courses/111/102/111102160/
2. https://www.geeksforgeeks.org/probability-and-statistics/
3. https://www.khanacademy.org/

युनिट - 4
इंटरपोलेशन
(Interpolation)

विषय निष्पत्ति (Course Outcome)

CO4 - इंटरपोलेशनवर आधारित योग्य पद्धत अंमलात आणा.

घटक निष्पत्ती (Theory learning outcomes)

1. लागरांगेस इंटरपोलेशन सूत्र वापरून समस्या सोडवा.
2. फॉरवर्ड आणि बॅकवर्ड डिफरन्स टेबल तयार करा.
3. फॉरवर्ड, बॅकवर्ड, शिफ्ट आणि इनव्हर्स शिफ्ट ऑपरेटर वापरून समस्या सोडवा.
4. फॉरवर्ड आणि बॅकवर्ड इंटरपोलेशनवरील समस्या सोडवा.
5. एक्सट्रॅपोलेशनवरील समस्या सोडवा.

4.1 परिचय (Introduction):

न्यूटनच्या पद्धतीचे सूत्र समीकरण अंदाजे सामने किंमत ठरविण्यासाठीं वापरले जाते. एका मुळापासून दुस-या मुळापर्यंतच्या पुनरावृत्तींद्वारे बहुपदीय समीकरणाच्या मुळांची गणना करण्यासाठी न्यूटनचे पद्धत सूत्र दिले आहे. दोन गणितज्ञ जेम्स ग्रेगरी आणि आयझॅक न्यूटन यांच्या नावावर सूत्रांची नावे ठेवण्यात आलेली आहेत, ज्यांनी 17 व्या शतकात स्वतंत्रपणे समान इंटरपोलेशन तंत्र विकसित केले.

येथे आपण खालील प्रमाणे तीन चार लेखांचा अभ्यास करणार आहोत

- लागरांगेस इंटरपोलेशन सूत्र
- फायनाईट डिफरेन्सस आणि शिफ्टिंग ऑपरेटर (E)
- ऑपरेटर Δ, ∇ आणि E यांच्यातील संबंध (Relations between the operators Δ, ∇ and E)
- न्यूटनचे फॉरवर्ड आणि बॅकवर्ड डिफरन्स
- एक्स्ट्रापोलेशनची संकल्पना (Concept of Extrapolation)

4.2 लागरांगेस इंटरपोलेशन सूत्र (Lagrange's interpolation formula):

न्यूटनचे फॉरवर्ड आणि बॅकवर्ड इंटरपोलेशन फॉर्म्युले केवळ तेव्हाच वापरले जाऊ शकतात जेव्हा x ची मूल्ये समान अंतरावर असतात. जर x ची मूल्ये समान अंतरावर असतील किंवा नसतील तर, आपण लागरांगेस (Lagrange) चे इंटरपोलेशन सूत्र वापरतो.

y =f(x) हे फंक्शन असू द्या की f(x) ही मूल्ये $y_0, y_1, y_2 .... y_n$ ला x= $x_0, x_1, x_2, ..., x_n$ला संबंधित आहेत. yi = f(xi),i = 0,1,2,...,n . आता, (n + 1) जोडलेली मूल्ये आहेत (xi, yi) i = 0, 1, 2,..., n आणि म्हणून f (x) हे x मधील अंश n च्या बहुपदी कार्याद्वारे दर्शविले जाऊ शकते, मग लागरांगेस चे सूत्र पुढील प्रमाणे आहे,

$$y = f(x) = \frac{(x-x_1)(x-x_2)\dots(x-x_n)}{(x_0-x_1)(x_0-x_2)\dots(x_0-x_n)}y_0 + \frac{(x-x_0)(x-x_2)\dots(x-x_n)}{(x_1-x_0)(x_1-x_2)\dots(x_1-x_n)}y_1 + \cdots. + \frac{(x-x_0)(x-x_1)\dots(x-x_{n-1})}{(x_n-x_0)(x_n-x_1)\dots(x_n-x_{n-1})}y_n$$

सोडवलेली उदाहरणे (Solved Problems)

उदाहरण 4.1: लागरांगेस च्या इंटरपोलेशन सूत्राचा वापर करून f(10) काढा.

X	5	6	9	11
y	12	13	14	16

उत्तर :

येथे मध्यांतर असमान आहेत. लागरांगेस इंटरपोलेशन सूत्र वापरा.

$x_0 = 5, x_1 = 6, x_2 = 9, x_3 = 11$

$y_0 = 12, y_1 = 13, y_2 = 14, y_3 = 16$ and x = 10

$$y=f(x)=\frac{(x-x_1)(x-x_2)\dots(x-x_n)}{(x_0-x_1)(x_0-x_2)\dots(x_0-x_n)}y_0 + \frac{(x-x_0)(x-x_2)\dots(x-x_n)}{(x_1-x_0)(x_1-x_2)\dots(x_1-x_n)}y_1 + \frac{(x-x_0)(x-x_1)(x-x_3)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)}y_2 + \frac{(x-x_0)(x-x_1)(x-x_2)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)}y_3$$

$$= \frac{(x-6)(x-9)(x-11)}{(5-6)(5-9)(5-11)} \times 12 + \frac{(x-5)(x-9)(x-11)}{(6-5)(6-9)(6-11)} \times 13 + \frac{(x-5)(x-6)(x-11)}{(9-5)(9-6)(9-11)} \times 14 + \frac{(x-5)(x-6)(x-9)}{(11-5)(11-6)(11-9)} \times 16$$

x=10 टाका

$$f(10)=\frac{(4)(1)(-1))}{(-1)(-4)(-6)} \times 12 + \frac{(5)(1)(-1)}{(1)(-3)(-5)} \times 13 + \frac{(5)(4)(-1)}{(4)(3)(-2)} \times 14 + \frac{(5)(4)(1)}{((6)(5)(2)} \times 16$$

$$f(10) = \frac{1}{6}(12) - \frac{13}{3} + \frac{5\times 14}{3\times 2} + \frac{4\times 16}{12}$$

f(10) = 14.6663

उदाहरण 4.2: लागरांगेस चे सूत्र वापरा आणि खालील माहिती वरून अंदाज लावा की उत्पन्न मिळवणाऱ्या कामगारांची संख्या दरमहा रु. 26 पेक्षा जास्त नाही.

उत्पन्न (₹) पेक्षा जास्त नाही	15	25	30	35
कामगारांची संख्या	36	40	45	48

उत्तर :

येथे मध्यांतर असमान आहेत. लागरांगेस इंटरपोलेशन सूत्र वापरा, लागरांगेस इंटरपोलेशन सूत्र आपल्याकडे आहे.

$x_0 = 15, x_1 = 20, x_2 = 30, x_3 = 35$

$y_0 = 36, y_1 = 40, y_2 = 45, y_3 = 48$ and x=26

$$y = f(x) = \frac{(x-x_1)(x-x_2)\dots(x-x_n)}{(x_0-x_1)(x_0-x_2)\dots(x_0-x_n)}y_0 + \frac{(x-x_0)(x-x_2)\dots(x-x_n)}{(x_1-x_0)(x_1-x_2)\dots(x_1-x_n)}y_1 + \frac{(x-x_0)(x-x_1)(x-x_3)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)}y_2 + \frac{(x-x_0)(x-x_1)(x-x_2)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)}y_3$$

$$y = \frac{(26-25)(26-30)(26-35)}{(15-25)(15-30)(15-35)} \times 36 + \frac{(26-15)(26-30)(26-35)}{(25-15)(25-30)(25-35)} \times 40 + \frac{(26-15)(26-25)(26-35)}{(30-15)(30-25)(30-35)} \times 45 + \frac{(26-15)(26-25)(26-30)}{(35-15)(35-25)(35-30)} \times 38$$

$$f(26) = \frac{(1)(-4)(-9)}{(-10)(-15)(-20)} \times 36 + \frac{(11)(-4)(-9)}{(10)(-5)(-10)} \times 40 + \frac{(11)(1)(-9)}{(15)(5)(-5)} \times 45 + \frac{(9)(1)(-4)}{(-20)(10)(5)} \times 38$$

$$f(26) = -\frac{(36)(36)}{(150)(20)} + \frac{(44)(9)(40)}{500} + \frac{(99)(45)}{(15)(25)} + \frac{(36)(38)}{(200)(15)}$$

$$f(26) = -\frac{1296}{3000} + \frac{15840}{500} + \frac{4455}{375} + \frac{1368}{1000}$$

$$f(26) = -0.432 + 31.68 + 11.88 - 1.368$$

$$f(26) = 41.76$$

त्यामुळे दरमहा २६ रुपयांपेक्षा जास्त उत्पन्न न मिळवणाऱ्या कामगारांची संख्या ४२ आहे

उदाहरण 4.3: लागरांगेस इंटरपोलेशन वापरून,,x = १५ असताना f(x) चे मूल्य शोधा

x	3	7	11	19
f(x)	42	43	47	60

उत्तर :

येथे मध्यांतर असमान आहेत. लागरांगेस इंटरपोलेशन सूत्र वापरा. लागरांगेस इंटरपोलेशन सूत्र आपल्याकडे आहे.

$x_0 = 3, x_1 = 7, x_2 = 11, x_3 = 19$

$y_0 = 42, y_1 = 43$,$y_2 = 47$,$y_3 = 60$ and x=15

y=f(x)=$\frac{(x-x_1)(x-x_2)...(x-x_n)}{(x_0-x_1)(x_0-x_2)...(x_0-x_n)}y_0 + \frac{(x-x_0)(x-x_2)...(x-x_n)}{(x_1-x_0)(x_1-x_2)...(x_1-x_n)}y_1 +$

$\frac{(x-x_0)(x-x_1)(x-x_3)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)}\ y_2 + \frac{(x-x_0)(x-x_1)(x-x_2)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)}\ y_3$

y=$\frac{(15-7)(15-11)(15-19)}{(3-7)(3-11)(3-19)} \times 42 + \frac{(15-3)(15-11)(15-19)}{(7-3)(7-11)(7-19)} \times 43 + \frac{(15-3)(15-7)(15-19)}{(11-3)(11-7)(11-19)} \times 47 +$

$\frac{(15-3)(15-7)(15-11)}{(19-3)(19-7)(19-11)} \times 60$

$f(15) = \frac{(8)(4)(-4)}{(-4)(-8)(-16)} \times 42 + \frac{(12)(4)(-4)}{(4)(-4)(-12)} \times 43 + \frac{(12)(8)(-4)}{(8)(4)(-8)} \times 47 + \frac{(12)(8)(4)}{(16)(12)(8)} \times 60$

$f(15) = \frac{42}{4} - 43 + \frac{12\times 47}{8} + \frac{240}{16}$

$f(15) = 10.5 - 43 + 70.5 + 15$

$f(15) = 53$

जेव्हा x=15, f(x) = 53 आहे.

उदाहरण 4.4: इंटरपोलेशन वापरून खालील माहिती वरून 1985 मध्ये केलेल्या व्यवसायाचा अंदाज लावा

वर्ष	1982	1983	1984	1986
व्यवसाय उत्पन्न (लाखांमध्ये)	150	235	365	525

उत्तर :

येथे मध्यांतर असमान आहेत. लागरांगेस इंटरपोलेशन सूत्र वापरा.

लागरांगेस इंटरपोलेशन सूत्र आपल्याकडे आहे.

$x_0 = 1982, x_1 = 1983, x_2 = 1984, x_3 = 1986$

$y_0 = 150, y_1 = 235$,$y_2 = 365$,$y_3 = 525$ and x=1985

$y = f(x) = \frac{(x-x_1)(x-x_2)...(x-x_n)}{(x_0-x_1)(x_0-x_2)...(x_0-x_n)}y_0 + \frac{(x-x_0)(x-x_2)...(x-x_n)}{(x_1-x_0)(x_1-x_2)...(x_1-x_n)}y_1 +$

$\frac{(x-x_0)(x-x_1)(x-x_3)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)}\ y_2 + \frac{(x-x_0)(x-x_1)(x-x_2)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)}\ y_3$

$y = \frac{(1985-1983)(1985-1984)(1985-1986)}{(1982-1983)(1982-1984)(1982-1986)} \times 150 + \frac{(1985-1982)(1985-1984)(1985-1986)}{(1983-1982)(1983-1984)(1983-1986)} \times 235 +$

$\frac{(1985-1982)(1985-1983)(1985-1986)}{(1984-1982)(1984-1983)(1984-1986)} \times 365 + \frac{(1985-1982)(1985-1983)(1985-1986)}{(1983-1982)(1983-1984)(1983-1986)} \times 525$

$y(1985) = \frac{(2)(1)(-1)}{(-1)(-2)(-4)} \times 150 + \frac{(3)(1)(-1)}{(1)(-1)(-3)} \times 235 + \frac{(3)(2)(-1)}{(2)(1)(-2)} \times 365 + \frac{(3)(2)(1)}{(4)(3)(2)} \times 525$

$y(1985) = 0.25 \times 150 + (-1) \times 235 + 1.5 \times 365 + 0.25 \times 525$

$y(1985) = 481.25$

1985 बिंदूवरील बहुपदीचे समाधान y(1985)=481.25 आहे.

उदाहरण 4.5: इंटरपोलेशन वापरून खालील माहिती वरून $\log_{10} 13$ मध्ये केलेल्या व्यवसायाचा अंदाज लावा.

$\text{Log}_{10} 10 = 1$, $\log_{10} 12 = 1.1, \log_{10} 15 = 1.2$ *and* $\log_{10} 20 = 1.3$

उत्तर :

येथे मध्यांतर असमान आहेत. लागरांगेस इंटरपोलेशन सूत्र वापरा.

लागरांगेस इंटरपोलेशन सूत्र आपल्याकडे आहे.

$x_0 = 190, x_1 = 12, x_2 = 15, x_3 = 20$

$y_0 = 1, y_1 = 1.1$,$y_2 = 1.2$,$y_3 = 1.3$ and x=13

$$y = f(x) = \frac{(x-x_1)(x-x_2)\ldots(x-x_n)}{(x_0-x_1)(x_0-x_2)\ldots(x_0-x_n)} y_0 + \frac{(x-x_0)(x-x_2)\ldots(x-x_n)}{(x_1-x_0)(x_1-x_2)\ldots(x_1-x_n)} y_1 +$$

$$\frac{(x-x_0)(x-x_1)(x-x_3)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)}\ y_2 + \frac{(x-x_0)(x-x_1)(x-x_2)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)}\ y_3$$

$$y = \frac{(13-12)(13-15)(13-20)}{(10-12)(10-15)(10-20)} \times 1 + \frac{(13-10)(13-12)(13-20)}{(12-10)(12-15)(12-20)} \times 1.1 +$$

$$\frac{(13-10)(13-12)(13-20)}{(15-10)(15-12)(15-20)} \times 1.2 \ + \ \frac{(13-10)(13-12)(13-15)}{(20-10)(20-12)(20-15)} \times 1.3$$

$$y(1985) = \frac{(1)(-2)(7)}{(-2)(-5)(-10)} \times 1 + \frac{(3)(-2)(-7)}{(2)(-3)(-8)} \times 1.1 + \frac{(3)(1)(-7)}{(5)(3)(-5)} \times 1.2 + \frac{(3)(-1)(-2)}{(10)(8)(5)} \times 1.3$$

$y(1985) = -0.14 + 0.9625 + 0.336 + 0.0195$

$y(1985) = 1.178$

$\log_{10} 13$ बिंदूवरील बहुपदीचे समाधान 1.178आहे.

सराव प्रश्नसंच (Exersice)

1. लागरांगेस इंटरपोलेशन सूत्र वापरून x=10 वर y शोधा

x	5	6	9	11
y	12	13	14	16

2. लागरांगेस इंटरपोलेशन सूत्र वापरून f(5) शोधा

x	1	2	3	4	7
y	2	4	8	16	28

3. लागरांगेस इंटरपोलेशन सूत्र वापरून f(1.5) शोधा f(0)=3 , f(1)=6 ,f(2)=11, f(3)=18.
4. लागरांगेस इंटरपोलेशन सूत्र वापरून बहुपद शोधा जे खाली दिलेली मूल्ये घेते

x	0	1	2	3
y	1	4	15	40

5 लागरांगेस इंटरपोलेशन सूत्र वापरून बहुपद शोधा जे खाली दिलेली मूल्ये घेते

x	0	1	2
y	1	4	6

4.3 फायनाईट डिफरेन्सस (Finite Differences)

$x_0, x_1, x_2, \ldots\ldots, x_n$आणि $y_0, y_1, y_2 \ldots. y_n$ वितर्कांचा विचार करा. Y =f (x) हे x चे कार्य असावे. आपण असे गृहीत धरू की x ची मूल्ये वाढत्या क्रमाने आहेत आणि तितकेच अंतर h बरोबर अंतरावर आहे. नंतर x ची मूल्ये x_0 , x_0+h , x_0+2h ,.... , x_0+nh अशी घेतली जाऊ शकतात आणि फंक्शन f (x_0), f (x_0 + h), f (x_0) ही मूल्ये गृहीत धरते. + 2h), , f (x_0 + nh) . येथे आपण फंक्शन y = f (x) च्या काही मर्यादित फरकांचा अभ्यास करू.

फॉरवर्ड डिफरेन्स ऑपरेटर Forward Difference Operator(Δ):

y = f(x) हे x चे दिलेले कार्य समजा. $y_0, y_1, y_2 .... y_n$ ही अनुक्रमे y ची मूल्ये = $x_0, x_1, x_2, ..., x_n$ असू द्या. नंतर $y_1 - y_0, y_2 - y_1, y_3 - y_2, ..., y_n - y_{n-1}$ यांना y फंक्शनचे पहिले (फॉरवर्ड) फरक म्हणतात. ते अनुक्रमे Δy_0, $\Delta y_1, \Delta y_2, ..., \Delta y_{n-1}$ द्वारे दर्शविले जातात.

(i.e) $\Delta y_0 = y_1 - y_0$, $\Delta y_1 = y_2 - y_1$, $\Delta y_2 = y_3 - y_2, ..., \Delta y_{n-1} = y_n - y_{n-1}$

सर्वसाधारणपणे, $\Delta y_n = y_{n+1} - y_n$, n = 0,1,2,3,...

चिन्ह Δ ला फॉरवर्ड डिफरन्स ऑपरेटर म्हणतात आणि डेल्टा म्हणून उच्चारले जाते.

फॉरवर्ड डिफरन्स ऑपरेटर Δ हे Df (x) = f (x + h) – f (x) म्हणून देखील परिभाषित केले जाऊ शकते, h हे अंतराचे समान अंतर आहे.

ऑपरेटरचे गुणधर्म Δ : Properties of the operator Δ :

गुणधर्म 1: जर c स्थिर असेल तर Δc = 0 (If c is a constant then Δc = 0)

पुरावा: f(x) = c समजा

∴ f (x + h) = c (जेथे 'h' हा फरकाचा मध्यांतर आहे)

Δf (x) = f (x + h) – f (x)

Δc = c – c = 0

गुणधर्म 2: Δ वितरणात्मक आहे (Δ is distributive)

i.e. Δ (f (x) + g (x)) = Δ f (x) + Δ g (x)

पुरावा: Δ[f (x) + g (x)] = [f (x + h) + g(x + h)] – [f (x)+ g (x)]

= f (x + h) + g(x + h) – f (x) – g (x)

= f (x + h) – f (x)+ g(x + h) – g (x)

= Δ f (x) + Δ g (x)

त्याचप्रमाणे आपण ते दाखवू शकतो Δ [f (x) – g (x)] = f (x) – Δ [g (x)]

सामान्यतः, Δ[f1 (x) + f 2 (x)...... + f n (x)] = Δ f1 (x)+ Δ f 2 (x) + ... + Δ f n (x)

गुणधर्म 3: जर c स्थिर असेल तर Δ c f (x) = c Δ f (x)

If c is a constant then Δ c f (x) = c Δ f (x)

पुरावा: Δ [c f (x)] = c f (x + h) – c f (x)

= c [f (x + h) – f (x)]

= c Δ f (x)

$\Delta^2 y_0$, $\Delta^2 y_1$, ..., $\Delta^2 y_n$ द्वारे दर्शविलेल्या पहिल्या फरकांच्या फरकांना द्वितीय फरक म्हणतात, जेथे

$\Delta^2 y_n = \Delta(\Delta y_n) = \Delta(y_{n+1} - y_n)$

$= \Delta y_{n+1} - \Delta y_n$............(1)

$\Delta^2 y_n = \Delta y_{n+1} - \Delta y_n, n = 0,1,2, ...$

$\Delta^2 y_0 = \Delta y_1 - \Delta y_0$

$\Delta^2 y_1 = \Delta y_2 - \Delta y_1 , ...$

त्याचप्रमाणे दुसऱ्या भेदांच्या फरकांना तिसरा फरक म्हणतात.

$\Delta^3 y_n = \Delta^2 y_{n+1} - \Delta^2 y_n , n = 0,1,2, ...$

विशेषतः

$\Delta^3 y_0 = \Delta^2 y_1 - \Delta^2 y_0$

$\Delta^3 y_1 = \Delta^2 y_2 - \Delta^2 y_1$

.

.

$\Delta^k y_n = \Delta^{k-1} y_{n+1} - \Delta^{k-1} y_n, n = 0,1,2, \ldots \ldots$

खाली दर्शविल्याप्रमाणे वरील फरक सारणीमध्ये प्रस्तुत करणे सोयीचे आहे.

फॉरवर्ड फरक सारणी:

बॅकवर्ड डिफरेन्स ऑपरेटर Backward Difference Operator (∇)

y = f(x) हे x चे दिलेले कार्य समजा. $y_0, y_1, y_2 \ldots . y_n$ही अनुक्रमे y ची मूल्ये = $x_0, x_1, x_2, \ldots \ldots, x_n$असू द्या

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$	$\Delta^5 y$
x_0	y_0					
		Δy_0				
x_1	y_1		$\Delta^2 y_0$			
		Δy_1		$\Delta^3 y_0$		
x_2	y_2		$\Delta^2 y_1$		$\Delta^4 y_0$	
		Δy_2		$\Delta^3 y_1$		$\Delta^5 y_0$
x_3	y_3		$\Delta^2 y_2$		$\Delta^4 y_1$	
		Δy_3		$\Delta^3 y_2$		
x_4	y_4		$\Delta^2 y_3$			
		Δy_4				
x_5	y_5					

$y_1 - y_0 = \nabla y_1$

$y_2 - y_1 = \nabla y_2$

$y_n - y_{n-1} = \nabla y_n$ त्यांना प्रथम (मागास) फरक म्हणतात.

ऑपरेटर ∇ ला बॅकवर्ड डिफरन्स ऑपरेटर म्हणतात आणि नेप्ला असे उच्चारले जाते.

$\nabla^2 y_0$, $\nabla^2 y_1$, …., $\nabla^2 y_n$ द्वारे दर्शविलेल्या पहिल्या फरकांच्या फरकांना द्वितीय फरक म्हणतात, जेथे

$\nabla^2 y_n = \nabla y_n - \nabla y_{n-1}$………….(1)

$\nabla^k y_n = \nabla^{k-1} y_n - \nabla^{k-1} y_{n-1}, n = 0,1,2, \ldots ..$

बॅकवर्ड फरक सारणी:

x	y	∇y	$\nabla^2 y$	$\nabla^3 y$	$\nabla^4 y$
x_0	y_0				
		∇y_1			
x_1	y_1		$\nabla^2 y_2$		
		∇y_2		$\nabla^3 y_3$	
x_2	y_2		$\nabla^2 y_3$		$\nabla^4 y_4$
		∇y_3		$\nabla^3 y_4$	
x_3	y_3		$\nabla^2 y_4$		
		∇y_4			
x_4	y_4				

बॅकवर्ड फरक देखील खालीलप्रमाणे परिभाषित केला जाऊ शकतो.

$\nabla f(x) = f(x) - f(x - h)$

प्रथम फरक: $\nabla f(x + h) = f(x + h) - f(x)$

$\nabla f(x + 2h) = f(x + 2h) - f(x + h)$,... h हे अंतराचे मध्यांतर आहे.

दुसरा फरक: $\nabla^2 f(x + h) = \nabla(\nabla f(x + h)) = \nabla(f(x + h) - f(x))$

$= \nabla f(x + h) - \nabla f(x)$

$\nabla^2 f(x + 2h) = \nabla f(x + 2h) - \nabla f(x + h)$

तिसरा फरक: $\nabla^3 f(x + h) = \nabla^2 f(x + h) - \nabla^2 f(x)$

$\nabla^3 f(x + 2h) = \nabla^2 f(x + 2h) - \nabla^2 f(x + h)$

येथे आपण ते लक्षात घेतो, $\nabla f(x + h) = f(x + h) - f(x) = \Delta f(x)$

$\nabla f(x + 2h) = f(x + 2h) - f(x + h) = \Delta f(x + h)$

$\nabla^2 f(x + 2h) = \nabla f(x + 2h) - \nabla f(x + h) = \Delta f(x + h) - \Delta f(x)$

$= \Delta^2 f(x)$

सामान्यतः, $\nabla^n f(x + nh) = \Delta^n f(x)$

शिफ्ट ऑपरेटर (Shift Operator E): E ला शिफ्टिंग ऑपरेटर म्हणतात. त्याला विस्थापन ऑपरेटर देखील म्हणतात.

y = f (x) हे x आणि x_0 , x_0 +h, x_0 +2h, x_0 +3h ..., x_0 +nh ही x ची अनुक्रमिक मूल्ये असू द्या. मग ऑपरेटर ई म्हणून परिभाषित केले आहे.

$E[f(x_0)] = f(x_0+h)$

$E[f(x_0 + h)] = f(x_0 + 2h)$, $E[f(x_0 + 2h)] = f(x_0 + 3h)$,..

$E[f(x_0 + (n - 1)h)] = f(x_0 + nh)$

$E[f(x)] = f(x + h)$, h हे अंतराचे (समान) मध्यांतर आहे

$E^2 f(x)$ म्हणजे ऑपरेटर "E", f (x) वर दोनदा लागू केला जातो.

(i.e) $E^2 f(x) = E[E f(x)] = E[f(x + h) = f(x + 2h)$

सामान्यतः

$E^n f(x) = f(x + nh)$ and $E^{-n} f(x) = f(x - nh)$

ऑपरेटर E चे गुणधर्म:

1. $E[f_1(x)+ f_2(x)+.... f_n(x)] = E f_1(x) + E f_2(x)+...+E f_n(x)$
2. $E[cf(x)] = cE[f(x)]$; c स्थिर आहे
3. $E^m [E^n f(x)] = E^n [(E^m f(x)) = E^{m+n} f(x)$
4. जर 'n' हा धन पूर्णांक असेल तर $E^n [E^{-n} (f(x))] = f(x)$

नोंद:

y = f(x) हे x चे दिलेले कार्य समजा. $y_0, y_1, y_2....y_n$ही अनुक्रमे y ची मूल्ये = $x_0, x_1, x_2, \ldots.., x_n$असू द्या . मग E ची व्याख्या देखील केली जाऊ शकते

$Ey_0 = y_1$, $Ey_1=y_2$, ..., $E y_{n-1}=y_n$

$E[Ey_0] = E(y_1) = y_2$ आणि सर्वसाधारणपणे $E^n y_0=y_n$

4.4 ऑपरेटर Δ, ∇ आणि E यांच्यातील संबंध (Relations between the operators Δ, ∇ and E)

1. $\Delta \equiv E - 1$

पुरावा:

Δ च्या व्याख्येवरून आपल्याला हे माहित आहे $\Delta f(x) = f(x + h) - f(x)$ and

$E[f(x)] = f(x + h)$

जेथे h हा फरकाचा मध्यांतर आहे.

$\Delta f(x) = f(x + h) - f(x)$

$\Delta f(x) = Ef(x) - f(x)$

$\Rightarrow \Delta f(x) = (E - 1) f(x)$

$\Delta \equiv E - 1$

$\therefore \quad E \equiv 1 + \Delta$

2. $E\Delta = \Delta E$

पुरावा:

$E(\Delta f(x)) = E[f(x+h) - f(x)]$

$= E f(x+h) - E f(x)$

$= f(x+2h) - f(x+h)$

$= \Delta f(x+h)$

$= \Delta E f(x)$

$\Delta E = \Delta E$

3. $\nabla = E - 1 / E = 1 - E^{-1}$

पुरावा:

$\nabla f(x) = f(x) - f(x-h)$

$= f(x) - E^{-1} f(x)$

$= (1 - E^{-1}) f(x)$

$\nabla = (1 - E^{-1})$

$\nabla \equiv 1 - 1/E$

$\therefore \nabla = [E - 1]/E$

4.5 न्यूटन फॉरवर्ड डिफरन्स (Newton Forward Difference): इंटरपोलेशन ही दिलेल्या स्वतंत्र डेटा सेटमधून एक साधी फंक्शन मिळवण्याची एक पद्धत आहे जसे की फंक्शन प्रदान केलेल्या डेटा पॉइंट्समधून जाते. हे दिलेल्या डेटामधील डेटा पॉइंट्स निर्धारित करण्यात मदत करते. खालील सूत्राला न्यूटनचे फॉरवर्ड इंटरपोलेशन सूत्र म्हणतात.

$$y(x) = y_0 + p\Delta y_0 + \frac{p(p-1)}{2!} . \Delta^2 y_0 + \frac{p(p-1)(p-2)}{3!} . \Delta^3 y_0 \text{ येथे}$$

$$h = x_1 - x_0$$

$$p = \frac{x - x_0}{h}$$

सोडवलेली उदाहरणे

उदाहरण 4.5.1: न्यूटन फॉरवर्ड डिफरेन्स सारणी वापरून f(2) ची अंदाजित किंमत काढा

X	1	3	5	7
F(x)	24	120	336	720

उत्तर:

न्यूटन फॉरवर्ड डिफरेन्स टेबल खालील प्रमाणे

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$
1	24			
		96		
3	120		120	
		216		48
5	336		168	
		384		
7	720			

$$h = x_1 - x_0 = 3 - 1 = 2$$

$$p = \frac{x - x_0}{h} = \frac{2-1}{2} = 0.5$$

न्यूटन फॉरवर्ड डिफरेन्स इंटरपोलेशन सूत्र

$$y(x) = y_0 + p\Delta y_0 + \frac{p(p-1)}{2!} . \Delta^2 y_0 + \frac{p(p-1)(p-2)}{3!} . \Delta^3 y_0$$

$$y(1) = 24 + 0.5 \times 96 + \frac{0.5(0.5-1)}{2!} \times 120 + \frac{0.5(0.5-1)(0.5-2)}{3!} \times 48$$

$$y(1) = 24 + 48 - 15 + 3$$

$$y(1) = 60$$

उदाहरण 4.5.2: न्यूटन फॉरवर्ड डिफरेन्स सारणी वापरून f(1896) ची अंदाजित किंमत काढा

X	1891	1901	1911	1921	1931
F(x)	46	66	81	93	101

उत्तर:

न्यूटन फॉरवर्ड डिफरेन्स टेबल खालील प्रमाणे

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
1891	46				
		20			
1901	66		-5		
		15		2	
1911	81		-3		-3
		12		-1	
1921	93		-4		
		8			
1931	101				

$$h = x_1 - x_0 = 1901 - 1891 = 10$$

$$p = \frac{x - x_0}{h} = \frac{1896 - 1891}{10} = 0.5$$

न्यूटन फॉरवर्ड डिफरेन्स इंटरपोलेशन सूत्र

$$y(x) = y_0 + p\Delta y_0 + \frac{p(p-1)}{2!}.\Delta^2 y_0 + \frac{p(p-1)(p-2)}{3!}.\Delta^3 y_0 + \frac{p(p-1)(p-2)(p-3)}{4!}.\Delta^4 y_0$$

$$y(1) = 46 + 0.5 \times 20 + \frac{0.5(0.5-1)}{2!} \times -5 + \frac{0.5(0.5-1)(0.5-2)}{3!} \times 2 + \frac{0.5(0.5-1)(0.5-2)(0.5-3)}{4!} \times -3$$

$$y(1) = 46 + 10 + 0.625 + 0.125 + 0.117$$

$$y(1) = 56.867$$

उदाहरण 4.5.3: न्यूटन फॉरवर्ड डिफरेन्स सारणी वापरून f(0.12) ची अंदाजित किंमत काढा

X	0.10	0.15	0.20	0.25	0.30
F(x)	0.1003	0.1511	0.2027	0.2553	0.3093

उत्तर:

न्यूटन फॉरवर्ड डिफरेन्स टेबल खालील प्रमाणे

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
0.10	0.1003				
		0.0508			
0.15	0.1511		0.0008		
		0.0516		0.0002	
0.20	0.2027		0.0001		0.0002
		0.0526		0.0004	
0.25	0.2553		0.0014		
		0.054			
0.30	0.3093				

$$h = x_1 - x_0 = 0.15 - 0.1 = 0.05$$

$$p = \frac{x - x_0}{h} = \frac{0.12 - 0.1}{0.05} = 0.4$$

न्यूटन फॉरवर्ड डिफरेन्स इंटरपोलेशन सूत्र

$$y(x) = y_0 + p\Delta y_0 + \frac{p(p-1)}{2!}.\Delta^2 y_0 + \frac{p(p-1)(p-2)}{3!}.\Delta^3 y_0 + \frac{p(p-1)(p-2)(p-3)}{4!}.\Delta^4 y_0$$

$$y(1) = 0.1003 + 0.4 \times 0.0508 + \frac{0.4(0.4-1)}{2!} \times 0.0008 + \frac{0.4(0.4-1)(0.4-2)}{3!} \times 0.002 + \frac{0.4(0.4-1)(0.4-2)(0.4-3)}{4!} \times 0.0002$$

$$y(1) = 0.1003 + 0.0203 + 0 + 0 + 0$$

$$y(1) = 0.1205$$

सराव प्रश्नसंच (Exersice) :

1) न्यूटन फॉरवर्ड डिफरेन्स सारणी वापरून y(22) ची अंदाजित किंमत काढा

X	20	25	30	35	40	45
y	354	332	291	260	231	204

2) न्यूटन फॉरवर्ड डिफरेन्स सारणी वापरून y(218) ची अंदाजित किंमत काढा

X	100	150	200	250	300	350	400
Y	10.63	13.03	15.04	16.81	18.42	19.90	21.27

3) खालील तक्त्यामध्ये न्यूटन फॉरवर्ड डिफरेन्स सारणी वापरून y(1.5) ची अंदाजित किंमत काढा.

X	1	2	3	4	5	6	7
Y	1	8	27	64	125	216	343

4) खालील तक्त्यामध्ये न्यूटन फॉरवर्ड डिफरेन्स सारणी वापरून y(40) ची अंदाजित किंमत काढा.

X	0°	30°	45°	60°
Y	0	8	1	1.732

न्यूटन बॅकवर्ड डिफरेन्स (Newton Backward Difference)

खालील सूत्राला न्यूटन बॅकवर्ड इंटरपोलेशन सूत्र म्हणतात.

$$y(x) = y_0 + p\nabla y_0 + \frac{p(p+1)}{2!}\,.\nabla^2 y_0 + \frac{p(p+1)(p+2)}{3!}\,.\nabla^3 y_0 + \frac{p(p+1)(p+2)(p+3)}{4!}\,.\nabla^4 y_0$$

येथे

$h = x_1 - x_0$

$p = \frac{x - x_n}{h}$

सोडवलेली उदाहरणे (Solved Problems)

उदाहरण 4.5.4: न्यूटन बॅकवर्ड डिफरेन्स सारणी वापरून f(1925) ची अंदाजित किंमत काढा

X	1891	1901	1911	1921	1931
F(x)	46	66	81	93	101

उत्तर:

न्यूटन बॅकवर्ड डिफरेन्स टेबल खालील प्रमाणे

x	y	∇y	$\nabla^2 y$	$\nabla^3 y$	$\nabla^4 y$
1891	46				
		20			
1901	66		-5		
		15		2	
1911	81		-3		-3
		12		-1	
1921	93		-4		
		8			
1931	101				

$h = x_1 - x_0 = 1901 - 1891 = 10$

$$p = \frac{x - x_n}{h} = \frac{1925 - 1931}{10} = -0.6$$

न्यूटन बॅकवर्ड डिफरेन्स इंटरपोलेशन सूत्र

$$y(x) = y_0 + p\nabla y_0 + \frac{p(p+1)}{2!}.\nabla^2 y_0 + \frac{p(p+1)(p+2)}{3!}.\nabla^3 y_0 + \frac{p(p+1)(p+2)(p+3)}{4!}.\nabla^4 y_0$$

$$y(1925) = 101 + (-0.6) \times 8 + \frac{-0.6(-0.6+1)}{2!} \times (-4) + \frac{-0.6(-0.6+1)(-0.6+2)}{3!} \times (-1) + \frac{-0.6(-0.6+1)(-0.6+2)(-0.6+3)}{4!} \times (-3)$$

$y(1925) = 101 - 4.8 + 0.056 + 0.1008$

$y(1925) = 96.8368$

उदाहरण 4.5.6: न्यूटन बॅकवर्ड डिफरेन्स सारणी वापरून f(4) ची अंदाजित किंमत काढा

X	0	1	2	3
F(x)	1	0	1	10

उत्तर:

न्यूटन बॅकवर्ड डिफरेन्स टेबल खालील प्रमाणे

x	y	∇y	$\nabla^2 y$	$\nabla^3 y$
0	1			
		-1		
1	0		2	
		1		6
2	1		8	
		9		
3	10			

$h = x_1 - x_0 = 1 - 0 = 1$

$$p = \frac{x - x_n}{h} = \frac{4 - 3}{1} = 1$$

न्यूटन बॅकवर्ड डिफरेन्स इंटरपोलेशन सूत्र

$$y(x) = y_0 + p\nabla y_0 + \frac{p(p+1)}{2!}.\nabla^2 y_0 + \frac{p(p+1)(p+2)}{3!}.\nabla^3 y_0$$

$$y(4) = 10 + 1 \times 9 + \frac{1(1+1)}{2!} \times 8 + \frac{1(1+1)(1+2)}{3!} \times 6$$

$y(4) = 10 + 9 + 8 + 6$

$y(4) = 33$

उदाहरण 4.5.7: न्यूटन बॅकवर्ड डिफरेन्स सारणी वापरून $\boldsymbol{x^3 - x + 1}$ समीकरणाचे f(3.75) ची किंमत शोधा x1 = 2 and x2 = 4 स्टेप व्हॅल्यु (h) = 0.5

उत्तर: $f(x) = x^3 - x + 1$ हे समीकरण आहे

x आणि y च्या किमतीचा तक्ता खालील प्रमाणे

x	y	∇y	$\nabla^2 y$	$\nabla^3 y$	$\nabla^4 y$
2	7				
		7.125			
2.5	14.125		3.75		
		10.875		0.75	
3	25		4.5		0
		15.375		0.75	
3.5	40.375		5.25		
		20.625			
4	61				

$$h = x_1 - x_0 = 2.5 - 2 = 0.5$$

$$p = \frac{x - x_n}{h} = \frac{3.75 - 4}{0.5} = -0.5$$

न्यूटन बॅकवर्ड डिफरेन्स इंटरपोलेशन सूत्र

$$f(x) = y_0 + p\nabla y_0 + \frac{p(p+1)}{2!} \,.\nabla^2 y_0 + \frac{p(p+1)(p+2)}{3!} \,.\nabla^3 y_0 + \frac{p(p+1)(p+2)(p+3)}{4!} \,.\nabla^4 y_0$$

$$y(3.75) = 61 + (-0.5) \times 20.625 + \frac{-0.5(-0.5+1)}{2!} \times 5.25 + \frac{-0.5(-0.5+1)(-0.5+2)}{3!} \times 0.75 + \frac{-0.5(-0.5+1)(-0.5+2)(-0.5+3)}{4!} \times 0$$

$$y(3.75) = 61 - 10.3125 - 0.65625 - 0.0468 + 0 = 49.9843$$

सराव प्रश्नसंच (Excersice):

1) न्यूटन बॅकवर्ड डिफरेन्स सारणी वापरून y(7.5) ची अंदाजित किंमत काढा

X	1	2	3	4	5	6	7	8
y	1	8	27	64	125	216	343	512

2) न्यूटन बॅकवर्ड डिफरेन्स सारणी वापरून y(375) ची अंदाजित किंमत काढा

X	100	150	200	250	300	350	400
Y	10.63	13.03	15.04	16.81	18.42	19.90	21.27

3) खालील तक्त्यामध्ये न्यूटन बॅकवर्ड डिफरेन्स सारणी वापरून y(3.5) ची अंदाजित किंमत काढा.

X	0	1	2	3	4
Y	7	17	45	103	203

4) खालील तक्त्यामध्ये न्यूटन बॅकवर्ड डिफरेन्स सारणी वापरून y(50) ची अंदाजित किंमत काढा.

X	0°	30°	45°	60°
Y	0	8	1	1.732

4.6 एक्स्ट्रापोलेशनची संकल्पना (Concept of Extrapolation)

y=f(x) हे एक फंकशन आहे. $y_0, y_1, y_2, ...$ह्या $x_0, x_1, x_2, ...$ची संबंधित मूल्ये आहेत. जर आपल्याला $[x_0, x_n]$ कोणत्याही बिंदूवर y चे मूल्य शोधायचे असेल तर आपण इंटरपोलेशन वापरतो परंतु मध्यांतराच्या बाहेर y चे मूल्य असेल तर आपण एक्सट्रापोलेशनं (Extrapolation) वापरतो.

जर हे मूल्य टेबलच्या सुरूवातीच्या जवळ असेल तर आपण न्यूटन फॉरवर्ड डिफरेन्स वापरतो जर हे मूल्य टेबलच्या शेवटच्या जवळ असेल तर आपण न्यूटन बॅकवर्ड डिफरेन्स वापरतो

सोडवलेली उदाहरणे (Solved Problems)

उदाहरण 4.5.8: इंटरपोलेशन वापरून f(9) ची अंदाजित किंमत काढा

X	0	2	4	6	8
f(x)	2	5	10	17	6

उत्तर: न्यूटन बॅकवर्ड डिफरेन्स टेबल खालील प्रमाणे

x	y	∇y	$\nabla^2 y$	$\nabla^3 y$	$\nabla^4 y$
0	2				
		3			
2	5		2		
		5		0	
4	10		2		0
		7		0	
6	17		2		
		9			
8	26				

$h = x_1 - x_0 = 2 - 0 = 2$

$p = \frac{x - x_n}{h} = \frac{9-8}{2} = 0.5$

न्यूटन बॅकवर्ड डिफरेन्स इंटरपोलेशन सूत्र

$$(x) = y_0 + p\nabla y_0 + \frac{p(p+1)}{2!}\,.\nabla^2 y_0 + \frac{p(p+1)(p+2)}{3!}\,.\nabla^3 y_0$$

$$y(9) = 26 + (0.5) \times 9 + \frac{0.5(0.5+1)}{2!} \times 2 + \frac{0.5(0.5+1)(0.5+2)}{3!} \times 0$$

$$y(3.75) = 26 + 4.5 + 0.75 + 0 = 31.25$$

उदाहरण 4.5.9: इंटरपोलेशन वापरून f(2) ची अंदाजित किंमत काढा

X	3	5	7	9
f(x)	9	25	49	81

उत्तर:

न्यूटन फॉरवर्ड डिफरेन्स टेबल खालील प्रमाणे

x	y	∇y	$\nabla^2 y$	$\nabla^3 y$
3	9			
		16		
5	25		8	
		24		0
7	49		8	
		32		
9	81			

$$h = x_1 - x_0 = 5 - 3 = 2$$

$$p = \frac{x - x_0}{h} = \frac{2 - 3}{2} = -0.5$$

न्यूटन फॉरवर्ड डिफरेन्स इंटरपोलेशन सूत्र

$$y(x) = y_0 + p\Delta y_0 + \frac{p(p-1)}{2!}.\Delta^2 y_0 + \frac{p(p-1)(p-2)}{3!}.\Delta^3 y_0$$

$$y(2) = 9 + (-0.5) \times 16 + \frac{-0.5(-0.5-1)}{2!} \times 8 + \frac{-0.5(-0.5-1)(-0.5-2)}{3!} \times 0$$

$$y(2) = 9 - 8 + 3 + 0$$

$$y(2) = 4$$

सराव प्रश्नसंच (Exercise):

1) सारणी वापरून y(7.5) ची अंदाजित किंमत काढा

X	1	2	3	4	5	6	7
y	1	8	27	64	125	216	343

2) y(0.4) ची अंदाजित किंमत काढा

X	1	2	3	4	5
Y	2	5	7	14	32

3) खालील सारणी वापरून y(5) ची अंदाजित किंमत काढा.

X	0	1	2	3	4
Y	-5	-10	-9	4	35

4) खालील सारणी वापरून y(22) ची अंदाजित किंमत काढा.

X	10	15	20
Y	14	18	28

संदर्भ (Reference):

1. https://byjus.com/lagrange-interpolation-formula/
2. http://www.mathsisfun.com/
3. http://tutorial.math.lamar.edu/

युनिट – 5
सॅम्पलिंग मेथोड
(Sampling Method)

विषय निश्पती :(**Course Outcome)**R प्रोग्रामिंग वापरून दिलेल्या समस्येवर सॅम्पलिंग पद्धत लागू करा
घटक निश्पती :(Theory Learning Outcome)
1. सॅम्पलिंग वापरून दिलेल्यासमस्येचे निराकरण करा
2. टी वितरण 9T Distribution) वापरून नमुने तपासा.
3. ची स्क्वेअर .वितरण वापरून नमुने तपासा (chi-square)
4. स्वतंत्रतेची चाचणी घेण्यासाठी ची स्क्वेअरचाचणी वापरा.

सॅम्पलिंग पद्धतीमध्ये डेटा संकलित करण्यासाठी आणि संपूर्ण लोकसंख्येबद्दल निष्कर्ष काढण्यासाठी मोठ्या लोकसंख्येतील व्यक्ती किंवा निरीक्षणांचा उपसंच निवडणे समाविष्ट आहेलोकसंख्येच्या प्रत्येक सदस्याकडून . माहिती गोळा करणे अव्यवहार्य किंवा अशक्य असताना डेटा गोळा करण्याचा हा एक व्यावहारिक आणि कार्यक्षम मार्ग आहे.

5.1.1 Population(लोकसंख्या): लोकसंख्या हा संपूर्ण गट आहे ज्याबद्दल तुम्हाला निष्कर्ष काढायचा आहे. लोकसंख्येमध्ये सर्व व्यक्ती, वस्तू किंवा अभ्यासाधीन निरीक्षणे समाविष्ट असताना, नमुना विश्लेषणासाठी निवडलेल्या उपसंचाचे प्रतिनिधित्व करतो. ही विषमता ओळखून संशोधकांना नमुन्याच्या वैशिष्ट्यांवर आधारित मोठ्या गटाबद्दल अर्थपूर्ण निष्कर्ष काढता येतात.
लोकसंख्येचे विविध प्रकार आहेत.

1) मर्यादित लोकसंख्या (Finite Population): मर्यादित लोकसंख्येला मोजण्यायोग्य (countable) लोकसंख्या म्हणून देखील ओळखले जाते ज्यामध्ये लोकसंख्या मोजली जाऊ शकते. दुसऱ्या शब्दांत, हे सर्व व्यक्ती किंवा वस्तूंची लोकसंख्या म्हणून परिभाषित केले जाते जे मर्यादित आहेत. सांख्यिकीय विश्लेषणासाठी, मर्यादित लोकसंख्या असीम लोकसंख्येपेक्षा अधिक फायदेशीर आहे. मर्यादित लोकसंख्येची उदाहरणे म्हणजे कंपनीचे कर्मचारी, बाजारातील संभाव्य ग्राहक.
उदाहरण:
समजा एका शाळेतील 100 विद्यार्थी आहेत, आणि तुम्ही त्यांच्या 5वी वर्गाच्या गणिताचे पेपर तपासणार आहात. इथे शाळेतील 100 विद्यार्थी एक सिमित लोकसंख्या आहे.तुम्ही यातील काही विद्यार्थ्यांचे मतदान किंवा सर्वेक्षण करणार असाल, तर यामध्ये तुम्हाला प्रत्येक विद्यार्थ्याची माहिती मिळवणे आवश्यक आहे, आणि लोकसंख्येची संख्या निश्चित आहे. त्याचप्रमाणे, सिमित लोकसंख्येमध्ये एक मर्यादा आहे, उदाहरणार्थ, एका गावात 500 घरं असू शकतात. या 500 घरांमध्ये तुम्ही शोध घेतला तरी लोकसंख्या निश्चित आहे आणि ती 500 पेक्षा जास्त होणार नाही.

2) अनंत लोकसंख्या (Infinite Population): असीम (infinite) लोकसंख्येला अगणित (uncountable) लोकसंख्या म्हणून देखील ओळखले जाते ज्यामध्ये लोकसंख्येतील एककांची गणना करणे शक्य नाही. असीम लोकसंख्येचे उदाहरण म्हणजे रुग्णाच्या शरीरातील जंतूंची संख्या अगणित आहे.
उदाहरण:
बाजारातील ग्राहक: मानलं की एका मोठ्या शहरात एका वीकली बाजारात अनेक लोक येतात. या लोकसंख्येची संख्या दर वेळी बदलत असते कारण बाजारात वेगवेगळ्या वेळात ग्राहक येतात आणि जातात. त्यामुळं, ह्या लोकसंख्येची संख्या "अनंत" मानली जाऊ शकते, कारण ग्राहकांच्या आगमनाची आणि जाण्याची प्रक्रिया सतत सुरू असते.

सर्व इंटरनेट वापरकर्ते: एक असं उदाहरण घेता येईल, जिथे इंटरनेट वापरकर्त्यांची लोकसंख्या अनंत असू शकते, कारण दररोज नवीन लोक इंटरनेटचा वापर सुरू करतात. याची संख्या अचूकपणे मोजता येत नाही, आणि हे सतत वाढतच राहते. तुम्ही अशा लोकसंख्येमध्ये काही सदस्यांचा अभ्यास करत असताना, तुम्हाला सर्व लोकसंख्येचा अचूक अभ्यास किंवा मोजणी करणे कठीण जाते.

3) विद्यमान लोकसंख्या (Existent Population): विद्यमान लोकसंख्या ठोस व्यक्तींची लोकसंख्या म्हणून परिभाषित केली जाते. दुसऱ्या शब्दांत सांगायचे तर, ज्या लोकसंख्येचे एकक घन स्वरूपात उपलब्ध आहे तिला अस्तित्वात असलेली लोकसंख्या म्हणून ओळखले जाते. उदाहरणे म्हणजे पुस्तके, विद्यार्थी इ.

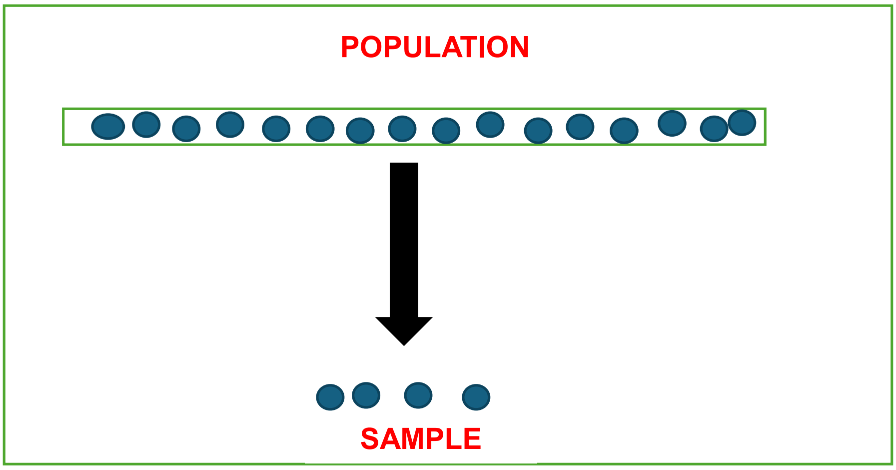
उदाहरण:

एका शाळेतील विदयार्थी: एका शाळेतील विदयार्थींची लोकसंख्या जी त्या विशिष्ट शाळेच्या एकाच शालेय वर्षात अस्तित्वात आहे, ती अस्तित्वात असलेली लोकसंख्या होईल. उदाहरणार्थ, शाळेतील 500 विद्यार्थी ज्यांचे शिक्षण चालू आहे, हे त्या शालेय वर्षात अस्तित्वात असलेल्या लोकसंख्येचे घटक आहेत. रुग्णालयात असलेले रुग्ण: समजा, एका रुग्णालयात 100 रुग्ण उपचार घेत आहेत. त्या रुग्णालयात असलेली लोकसंख्या म्हणजे त्या विशिष्ट वेळेत रुग्णालयात असलेले 100 रुग्ण, जे त्या कालावधीत अस्तित्वात आहेत. अशा लोकसंख्येमध्ये, आपण ज्या ठिकाणी आणि ज्या वेळेस लोकसंख्येचा अभ्यास करत आहोत, त्याच लोकसंख्येचा संदर्भ घेतला जातो.

4)काल्पनिक लोकसंख्या (Hypothetical Population): ज्या लोकसंख्येमध्ये एकक घन स्वरूपात उपलब्ध नाही तिला काल्पनिक लोकसंख्या असे म्हणतात. लोकसंख्येमध्ये निरीक्षणांचे संच, वस्तू इत्यादी असतात जे सर्व काही समान असतात. काही परिस्थितींमध्ये, लोकसंख्या केवळ काल्पनिक असते. उदाहरणे म्हणजे फासे गुंडाळण्याचा परिणाम, नाणे फेकण्याचा परिणाम.

उदाहरण:

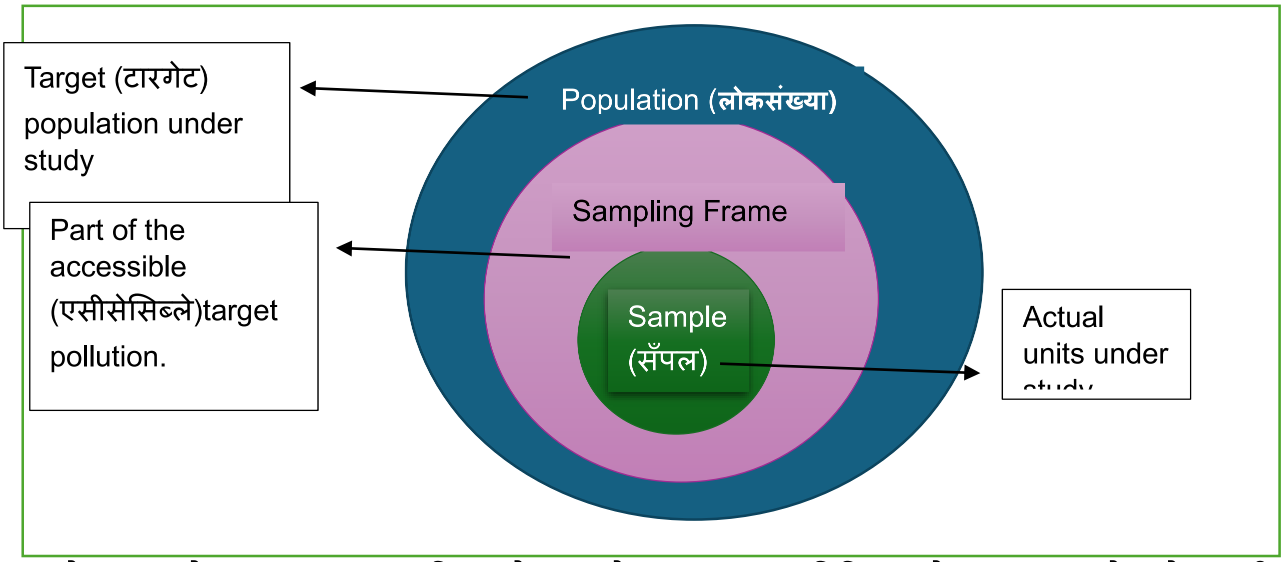
तुम्ही एका सर्वेक्षणात असे गृहित धरता की एका शहरात सर्व लोक 30-40 वयोगटातील आहेत. वास्तविकतेत, त्या शहरात प्रत्येक वयोगटातील लोक असू शकतात, पण तुमच्या संशोधनासाठी काल्पनिक लोकसंख्या तयार केली जाते, जिच्यात फक्त 30-40 वयोगटातील लोक असतील. यामध्ये, तुम्ही विचार करत असलेली लोकसंख्या वास्तविक लोकसंख्येची परिकल्पना आहे. एका गणितीय मॉडेलमध्ये कधीकधी "सर्व लोक समोच्च आहेत" असे गृहित धरून लोकसंख्या तयार केली जाते. या मॉडेलमध्ये, लोकांची उंची, वय, किंवा इतर शारीरिक लक्षणे काल्पनिक असतात, कारण हा फक्त एक गृहितक आहे. वैयक्तिक निवडण्याच्या समस्येत (Sampling Problem), समजा तुम्ही एक काल्पनिक लोकसंख्या तयार करता जी फक्त 1000 लोकांची असेल, आणि त्यात वेगवेगळ्या वयाच्या, लिंगाच्या, आणि इतर गुणधर्मांच्या लोकांची विविधता असते. हा एक "काल्पनिक" सेट असेल जो वास्तविक लोकसंख्येचा प्रतिनिधित्व करणारा असेल.काल्पनिक लोकसंख्या ही थोडक्यात एक गृहितक असते, जी वास्तविक परिस्थितीच्या आधारावर तयार केली जाते, पण ती प्रत्यक्षात अस्तित्वात नसते.



5.1.2 Sampling(सॅम्पलिंग): सॅम्पलिंग संशोधकांना लोकसंख्येच्या प्रतिनिधी उपसंचातून डेटा संकलित करण्यास अनुमती देते. लोकसंख्येच्या प्रत्येक सदस्याकडून डेटा संकलित करणे हे सहसा अव्यवहार्य(impractical) किंवा अशक्य असते, म्हणून सॅम्पलिंग संशोधकांना लोकसंख्येच्या लहान, अधिक व्यवस्थापित करण्यायोग्य उपसंचातून डेटा गोळा करण्यास अनुमती देते. नमुना हा लोकसंख्येमधून घेतलेला लहान गट आहे. नमुना हा घटकांचा समूह आहे ज्यातून तुम्ही प्रत्यक्षात डेटा गोळा कराल.

5.1.3 सॅम्पलिंगचे उद्दिष्ट (Aim of sampling):

मोठ्या लोकसंख्येच्या संशोधन प्रश्नाशी संबंधित असलेल्या वैशिष्ट्यांचे अंदाजे अंदाज घेणे हे सॅम्पलिंगचे उद्दिष्ट आहे. नमुना प्रातिनिधिक असणे आवश्यक आहे जेणेकरून संशोधक मोठ्या लोकसंख्येबद्दल निष्कर्ष काढू शकतील.



नमुना घेण्याचा उद्देश एक लहान संच मिळवणे हा आहे ज्यावर आपण विशिष्ट प्रयोग करू शकतो असे वाजवी गृहीत धरून की लहान संच प्रत्यक्षात लोकसंख्येची वैशिष्ट्ये प्रतिबिंबित करतो आणि या नमुन्यावर पाहिलेली कोणतीही निरीक्षणे मोठ्या लोकसंख्येला लागू होतात.

5.1.4 पॅरामीटर आणि सांख्यिकी(Parameter And Statistics): पॅरामीटर ही संपूर्ण लोकसंख्येचे वर्णन करणारी संख्या आहे (उदा. लोकसंख्येचे सरासरी), तर आकडेवारी म्हणजे नमुन्याचे वर्णन करणारी संख्या (उदा. नमुना सरासरी).

मापदंड शोधून(quantitative research) लोकसंख्येची वैशिष्ट्ये समजून घेणे हे परिमाणात्मक संशोधनाचे उद्दिष्ट आहे. व्यवहारात, लोकसंख्येच्या प्रत्येक सदस्याकडून डेटा संकलित करणे बऱ्याचदा कठीण, वेळखाऊ किंवा अव्यवहार्य असते. त्याऐवजी, नमुन्यांमधून डेटा गोळा केला जातो.

आकडेवारी वि पॅरामीटर्सची उदाहरणे (Examples of statistics vs parameters):

नमुना सांख्यिकी	**लोकसंख्या पॅरामीटर**
मृत्युदंडाचे समर्थन करणाऱ्या 2000 यादृच्छिकपणे(randomly) नमुना घेतलेल्या सहभागींचे प्रमाण.	मृत्युदंडाचे समर्थन करणाऱ्या सर्व यूएस रहिवाशांचे प्रमाण
बोस्टन आणि वेलस्लीमधील 850 महाविद्यालयीन विद्यार्थ्यांचे सरासरी उत्पन्न.	मॅसॅच्युसेट्समधील सर्व महाविद्यालयीन विद्यार्थ्यांचे सरासरी उत्पन्न.

एका शेतातील एवोकॅडोच्या(avocados) वजनाचे मानक विचलन(Standard deviation)	प्रदेशातील सर्व avocados (एवोकॅडोच्या) वजनाचे मानक विचलन(Standard deviation)
भारतातील 3000 हायस्कूल विद्यार्थ्यांचा सरासरी(mean) स्क्रीन वेळ.	भारतातील सर्व हायस्कूल विद्यार्थ्यांचा सरासरी(mean) स्क्रीन वेळ.

मापदंड आणि आकडेवारी ची तुलना (comparison of parameters and statistics):

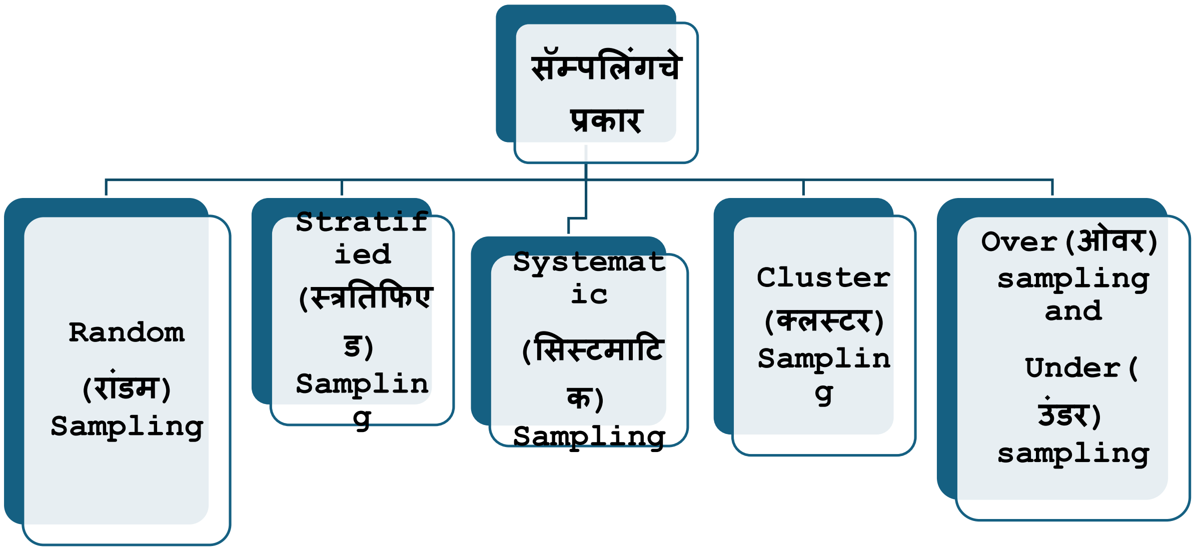
	पॅरामीटर	**सांख्यिकी**
व्याख्या	एक संख्यात्मक वैशिष्ट्य जे संपूर्ण लोकसंख्येचे वर्णन करते.	लोकसंख्येमधून काढलेल्या नमुन्यावरून मोजलेले संख्यात्मक माप.
निसर्ग(Nature)	स्थिर आणि स्थिर	चल (variable)आणि नमुना ते नमुन्यात बदलू शकतो
मूल्य(Value)	अज्ञात(unknown) (जसे ते संपूर्ण लोकसंख्येशी संबंधित आहे)	ज्ञात(known))(जसे ते नमुना डेटावरून घेतले आहे)
प्रतिनिधित्व (**Representation**)	संपूर्ण लोकसंख्येचे प्रतिनिधित्व	लोकसंख्येचा नमुना दर्शवते
वापर(Usages)	सैद्धांतिक संकल्पना आणि गृहितक सूत्रीकरण मध्ये वापरले	व्यावहारिक डेटा विश्लेषण आणि संशोधन वापरले
अचूकता	अधिक अचूक कारण ती संपूर्ण लोकसंख्येचे प्रतिनिधित्व करते	ते अचूकतेमध्ये भिन्न असू शकते कारण ते नमुना आकार आणि निवडीवर अवलंबून असते
गणना (Calculation)	बहुतेक वास्तविक-जगातील परिस्थितींमध्ये अचूकपणे गणना केली जाऊ शकत नाही	नमुना डेटा वापरून गणना केली जाऊ शकते
उद्देश	लोकसंख्येच्या वैशिष्ट्यांचे वर्णन करण्यासाठी	नमुना डेटावर आधारित लोकसंख्येच्या मापदंडांचा अंदाज लावणे
उदाहरण	देशातील सर्व प्रौढांची वास्तविक सरासरी उंची	देशातील प्रौढांच्या नमुना गटातून मोजलेली सरासरी उंची

5.1.5. सॅम्पलिंगचे प्रकार (Types of Sampling):

सॅम्पलिंग पद्धतीचे दोन प्राथमिक प्रकार आहेत

संभाव्यता सॅम्पलिंग(Probability sampling):संभाव्यता सॅम्पलिंगमध्ये यादृच्छिक(random) निवड समाविष्ट असते, ज्यामुळे तुम्हाला संपूर्ण गटाबद्दल मजबूत सांख्यिकीय निष्कर्ष काढता येतात.

गैर-संभाव्यता सॅम्पलिंग (Non-probability sampling):गैर-संभाव्यता सॅम्पलिंगमध्ये सोयी किंवा इतर निकषांवर आधारित नॉन-यादृच्छिक निवड समाविष्ट असते, ज्यामुळे तुम्हाला डेटा सहज गोळा करता येतो. संभाव्यता सॅम्पलिंग नमुन्याचे प्रकार खालीलप्रमाणे आहेत.



मशीन लर्निंगमधील सॅम्पलिंग पद्धतींचे विविध प्रकार, त्यांची वर्णने, पायथन कोड उदाहरणे आणि वापर प्रकरणांमध्ये विभागू या.

1.Random Sampling(रॅंडम सॅम्पलिंग): रॅंडम सॅम्पलिंग हा मशीन लर्निंगमधील सॅम्पलिंगचा सर्वात सोपा प्रकार आहे. यात कोणत्याही विशिष्ट पॅटर्नशिवाय डेटासेटमधून यादृच्छिकपणे डेटा पॉइंट निवडणे समाविष्ट आहे. ही पद्धत असे गृहीत धरते की डेटासेटमधील प्रत्येक डेटा पॉइंटला निवडण्याची समान संधी आहे.

Python Code (पायथन कोड) Example:

उदाहरण: लिस्टमधून रॅंडम सॅम्पल घेणे

python

Copy code

```
import random
# डेटा लिस्ट
data = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
# रॅंडम सॅम्पल (कुल ३ घटक निवडा)
sample = random.sample(data, 3)
print("रॅंडम सॅम्पल:", sample)
```

random.sample(population, k) हा फंक्शन तुम्हाला दिलेल्या population लिस्टमधून k आकाराचा रॅंडम सॅम्पल निवडतो. येथे k=3 म्हणजे ३ घटक निवडले जातील.

उदाहरण: रिप्लेसमेंटसह रॅंडम सॅम्पल घेणे

python

Copy code

```
import random
# डेटा लिस्ट
data = [1, 2, 3, 4, 5]
```

# रँडम सॅम्पल (रिप्लेसमेंटसह ५ घटक निवडा)
sample_with_replacement = random.choices(data, k=5)
print("रँडम सॅम्पल रिप्लेसमेंटसह:", sample_with_replacement)
random.choices(population, k) हा फंक्शन रिप्लेसमेंटसह रँडम सॅम्पल घेण्यासाठी वापरला जातो, यामध्ये एखादाच घटक पुन्हा निवडला जाऊ शकतो.
2. NumPy चा वापर:
जर तुम्ही मोठ्या डेटासेटवर काम करत असाल, तर numpy लायब्ररीचा वापर रँडम सॅम्पलिंगसाठी केला जातो.
उदाहरण: NumPy अ‍ॅरेमधून रँडम सॅम्पल घेणे
python
Copy code
```
import numpy as np
# NumPy अ‍ॅरे
data = np.array([1, 2, 3, 4, 5, 6, 7, 8, 9, 10])
# रँडम सॅम्पल (कुल ३ घटक निवडा, रिप्लेसमेंट शिवाय)
sample = np.random.choice(data, size=3, replace=False)
print("रँडम सॅम्पल (NumPy):", sample)
```
np.random.choice(population, size, replace) हा फंक्शन तुम्हाला NumPy अ‍ॅरेमधून रँडम सॅम्पल घेण्यासाठी वापरता येतो.
replace=False म्हणजे रिप्लेसमेंट नाही.
उदाहरण: NumPy अ‍ॅरेमधून रिप्लेसमेंटसह रँडम सॅम्पल
python
Copy code
```
import numpy as np
# NumPy अ‍ॅरे
data = np.array([1, 2, 3, 4, 5])
# रँडम सॅम्पल (रिप्लेसमेंटसह ५ घटक निवडा)
sample_with_replacement = np.random.choice(data, size=5, replace=True)
print("रँडम सॅम्पल रिप्लेसमेंटसह (NumPy):", sample_with_replacement)
```
3. DataFrame मध्ये रँडम सॅम्पल (pandas वापरून)
जर तुम्ही pandas DataFrame मध्ये काम करत असाल, तर तुम्ही sample() फंक्शनचा वापर करू शकता.
उदाहरण: DataFrame मध्ये रँडम सॅम्पल
python
Copy code
```
import pandas as pd
# DataFrame तयार करणे
df = pd.DataFrame({
    'A': [1, 2, 3, 4, 5],
    'B': ['a', 'b', 'c', 'd', 'e']
```

})
# रँडम सॅम्पल (२ पंक्ती निवडा)
sample_df = df.sample(n=2, replace=False)
print("DataFrame मधून रँडम सॅम्पल:\n", sample_df)

df.sample(n, replace) या पद्धतीने तुम्ही DataFrame मधून रँडम सॅम्पल निवडू शकता.

सारांश:

random.sample() हे साध्या लिस्टसाठी वापरता येते, जेथे रिप्लेसमेंट न करता रँडम सॅम्पल घ्यायचा असतो.

random.choices() वापरून रिप्लेसमेंटसह सॅम्पल घेतला जातो.

numpy.random.choice() हे NumPy अ‍ॅरेसाठी अधिक कार्यक्षम आणि लवचिक पद्धत आहे.

pandas.DataFrame.sample() वापरून DataFrame मधून रँडम सॅम्पल घ्या.

केसेस वापरा(Use Cases): यादृच्छिक(Random) नमुने सहसा प्रारंभिक अन्वेषण(exploratory) डेटा विश्लेषणामध्ये कोणत्याही पूर्वाग्रहाशिवाय डेटाचा द्रुत अर्थ प्राप्त करण्यासाठी वापरला जातो.

2. Stratified (स्त्रतिफिएड) Sampling: जेव्हा डेटासेटमध्ये उपसमूह असतात आणि प्रत्येक उपसमूहातील नमुने समाविष्ट करणे महत्त्वाचे असते तेव्हा स्तरीकृत नमुना वापरला जातो. या पद्धतीमध्ये, लोकसंख्या एकसंध उपसमूहांमध्ये विभागली जाते ज्याला स्तर म्हणतात, आणि नमुना संपूर्ण लोकसंख्येचा प्रतिनिधी आहे याची हमी देण्यासाठी प्रत्येक स्तरावरून योग्य संख्येचा नमुना घेतला जातो.

Python Code (पायथन कोड) Example:

Python मध्ये स्ट्रेटिफाइड सॅम्पलिंग कसा करायचा:

पायथनमध्ये स्ट्रेटिफाइड सॅम्पलिंग करण्यासाठी pandas लायब्ररीचा उपयोग केला जाऊ शकतो. इथे एक उदाहरण दिले आहे ज्यामध्ये डेटा फ्रेममध्ये असलेल्या विविध गटांमधून स्ट्रेटिफाइड सॅम्पल निवडले जात आहे.

उदाहरण:

python

Copy code

```
import pandas as pd
import numpy as np
# डेटा तयार करा
data = {
    'Age': [25, 30, 35, 40, 45, 50, 55, 60, 65, 70],
    'Gender': ['Male', 'Female', 'Male', 'Female', 'Male', 'Female', 'Male', 'Female', 'Male', 'Female'],
    'Income': [25000, 30000, 35000, 40000, 45000, 50000, 55000, 60000, 65000, 70000]
}
# DataFrame तयार करा
df = pd.DataFrame(data)
# 'Gender' या स्तंभानुसार स्ट्रेटिफाइड सॅम्पलिंग
stratified_sample = df.groupby('Gender', group_keys=False).apply(lambda x: x.sample(frac=0.5))
```

print("स्ट्रेटिफाइड सॅम्पलिंग:\n", stratified_sample)
सारांश :
groupby() पद्धतीचा वापर करून, डेटा फ्रेम 'Gender' स्तंभाच्या आधारावर गटांमध्ये विभागला जातो.
नंतर, प्रत्येक गटामधून sample(frac=0.5) या फंक्शनने ५०% डेटा निवडला जातो.
frac=0.5 म्हणजे प्रत्येक गटाच्या ५०% डेटा निवडा. तुम्ही frac बदलून आणखी कमी किंवा जास्त डेटा निवडू शकता.
group_keys=False हे सुनिश्चित करते की गटाच्या key (यात 'Gender') ला डेटामध्ये समाविष्ट केले जात नाही.
आउटपुट:
हा कोड काढल्यावर तुम्हाला प्रत्येक लिंगाच्या गटात ५०% रँडम सॅम्पल दिसेल. उदाहरणार्थ:
markdown
Copy code
स्ट्रेटिफाइड सॅम्पलिंग:

	Age	Gender	Income
0	25	Male	25000
3	40	Female	40000
1	30	Female	30000
4	45	Male	45000

अधिक गटांसह स्ट्रेटिफाइड सॅम्पलिंग:

स्ट्रेटिफाइड सॅम्पलिंग कोणत्याही गटावर आधारित करता येईल, जसे वय (Age), लिंग (Gender), इत्यादी. खाली एक उदाहरण आहे जिथे गट 'Age' वर आधारित आहे:

python
Copy code
```
import pandas as pd
import numpy as np
# डेटा तयार करा
data = {
    'Age': [25, 30, 35, 40, 45, 50, 55, 60, 65, 70],
    'Gender': ['Male', 'Female', 'Male', 'Female', 'Male', 'Female', 'Male', 'Female', 'Male', 'Female'],
    'Income': [25000, 30000, 35000, 40000, 45000, 50000, 55000, 60000, 65000, 70000]
}
# DataFrame तयार करा
df = pd.DataFrame(data)
# 'Age' गटावर आधारित स्ट्रेटिफाइड सॅम्पलिंग (उदा. 30-40 वयोगट)
age_bins = [20, 30, 40, 50, 60, 70]  # वयाचे गट
df['Age Group'] = pd.cut(df['Age'], bins=age_bins)
```

# प्रत्येक वयो गटातून सॅम्पल घ्या (५०% प्रत्येक गटातील)
stratified_sample = df.groupby('Age Group', group_keys=False).apply(lambda x: x.sample(frac=0.5))
print("स्ट्रेटिफाइड सॅम्पलिंग वयो गटावर आधारित:\n", stratified_sample)
आउटपुट:
markdown
Copy code
स्ट्रेटिफाइड सॅम्पलिंग वयो गटावर आधारित:

	Age	Gender	Income	Age Group
1	30	Female	30000	(20, 30]
0	25	Male	25000	(20, 30]
3	40	Female	40000	(30, 40]
5	50	Female	50000	(40, 50]
8	65	Male	65000	(60, 70]
6	55	Male	55000	(50, 60]

महत्त्वाचे मुद्दे:
pd.cut(): ह्याचा वापर करून तुम्ही वयोमानानुसार गट तयार करू शकता (उदाहरणार्थ, २०-३०, ३०-४० वयाचे गट).
groupby(): हे पद्धत गटांमध्ये विभागणी करते.
sample(frac=0.5): प्रत्येक गटापासून ५०% सॅम्पल घेते.
निष्कर्ष:
स्ट्रेटिफाइड सॅम्पलिंग मध्ये, तुम्ही विशिष्ट गटांपासून डेटा निवडता. त्यामुळे प्रत्येक गटाचे प्रतिनिधित्व चांगले होते.
pandas मध्ये groupby() आणि sample() चा वापर करून स्ट्रेटिफाइड सॅम्पलिंग खूप सोपे होते.
केसेस वापरा(Use Cases): फसवणूक शोधणे किंवा वैद्यकीय निदान यांसारख्या असमतोल डेटासेटशी व्यवहार करताना स्तरीकृत नमुने घेणे महत्त्वाचे असते, जेथे प्रत्येक वर्गाचे प्रतिनिधित्व महत्त्वाचे असते.
3. Systematic (सिस्टिमॅटिक) Sampling: सिस्टिमॅटिक सॅम्पलिंग हा सॅम्पलिंगचा एक प्रकार आहे जो एका निश्चित मध्यांतरानुसार डेटा पॉइंट्स निवडतो, ज्याला सॅम्पलिंग इंटरव्हल म्हणतात. उदाहरणार्थ, तुम्ही डेटासेटमधील प्रत्येक 10वा डेटा पॉइंट निवडू शकता.
Python Code (पायथन कोड) Example:
मजा आपल्याला एका लिस्टमधून प्रत्येक 3-वा सदस्य निवडायचा आहे, तर आपण खालील कोड वापरू शकतो:
python
Copy code
import numpy as np

# डेटा लिस्ट
data = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
# सॅम्पलिंग अंतर (k)
k = 3
# प्रारंभिक पॉइंट (म्हणजे कोणते घटक निवडायचे, उदाहरणार्थ 1-वा घटक)
start_point = 0
# सिस्टमॅटिक सॅम्पलिंग
sample = data[start_point::k]
print("सिस्टमॅटिक सॅम्पलिंग:", sample)
सारांश:
data[start_point::k]: हा कोड आपल्याला डेटा लिस्टमधून start_point पासून सुरुवात करून प्रत्येक k अंतरावर असलेल्या घटकांना निवडतो.
start_point=0 म्हणजे आपण लिस्टच्या पहिल्या घटकापासून सुरुवात करत आहोत.
k=3 म्हणजे प्रत्येक ३-वा घटक निवडला जातो.
आउटपुट:
less
Copy code
सिस्टमॅटिक सॅम्पलिंग: [1, 4, 7, 10]
2. सिस्टमॅटिक सॅम्पलिंग - DataFrame मध्ये उदाहरण:
आता समजा तुमच्याकडे pandas DataFrame आहे, आणि तुम्हाला त्यातील प्रत्येक 5-वा रेकॉर्ड निवडायचा आहे. आपण iloc किंवा iloc सॅलिंग पद्धतीचा वापर करू शकतो.
python
Copy code

```
import pandas as pd
# DataFrame तयार करा
data = {
    'ID': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10],
    'Name': ['A', 'B', 'C', 'D', 'E', 'F', 'G', 'H', 'I', 'J']
}
df = pd.DataFrame(data)
# सिस्टमॅटिक सॅम्पलिंग - प्रत्येक 5-वा रेकॉर्ड निवडायचा
k = 5  # सॅम्पलिंग अंतर
start_point = 0  # प्रारंभिक पॉइंट
# सिस्टमॅटिक सॅम्पलिंग DataFrame मध्ये
sample_df = df.iloc[start_point::k]
print("Static Sampling DataFrame:\n", sample_df)
```

काय होते इथे:
df.iloc[start_point::k]: या कोडने start_point=0 पासून सुरू करून प्रत्येक ५-वा रेकॉर्ड निवडला जातो.

k=5 म्हणजे प्रत्येक पाचव्या रेकॉर्डला निवडले जाते.

आउटपुट:

less

Copy code

सिस्टमॅटिक सॅम्पलिंग DataFrame मध्ये:

	ID	Name
0	1	A
5	6	F

3. सिस्टमॅटिक सॅम्पलिंग - शंभर रेकॉर्ड असलेल्या मोठ्या डेटावर उदाहरण:

कधी कधी आपल्याला मोठ्या डेटावर (उदाहरणार्थ 100 किंवा 1000 रेकॉर्ड) सिस्टमॅटिक सॅम्पलिंग करायचे असते. त्यासाठी खालीलप्रमाणे कोड असेल:

python

Copy code

```
import pandas as pd
# मोठ्या डेटा सेटचे उदाहरण (100 रेकॉर्ड)
data = {
    'ID': range(1, 101),
    'Name': ['Name_' + str(i) for i in range(1, 101)]
}
df = pd.DataFrame(data)
# सिस्टमॅटिक सॅम्पलिंग - प्रत्येक 10-वा रेकॉर्ड निवडायचा
k = 10  # सॅम्पलिंग अंतर
start_point = 0  # प्रारंभिक पॉइंट
# सिस्टमॅटिक सॅम्पलिंग DataFrame मध्ये
sample_df = df.iloc[start_point::k]
print("Systematic DataFrame (10th Record):\n", sample_df)
```

आउटपुट:

python

Copy code

सिस्टमॅटिक सॅम्पलिंग DataFrame मध्ये (१०-वा रेकॉर्ड):

	ID	Name
0	1	Name_1
9	10	Name_10
19	20	Name_20
29	30	Name_30
39	40	Name_40

निष्कर्ष:
सिस्टमॅटिक सॅम्पलिंग मध्ये तुम्ही डेटा लिस्ट किंवा डेटा फ्रेममधून निश्चित अंतरावरून घटक निवडता.
तुम्ही start_point आणि k (सॅम्पलिंग अंतर) बदलून तुम्ही कोणतेही इतर रेकॉर्ड निवडू शकता.
केसेस वापरा(Use Cases): तुमच्याकडे मोठा डेटासेट असेल आणि तुम्हाला संपूर्ण डेटासेटमध्ये पसरलेला नमुना हवा असेल तेव्हा पद्धतशीर नमुना घेणे उपयुक्त ठरते.

4. Cluster (क्लस्टर) Sampling: क्लस्टर सॅम्पलिंगमध्ये लोकसंख्येचे क्लस्टरमध्ये विभाजन करणे, त्यानंतर यादृच्छिकपणे (randomly)) संपूर्ण क्लस्टर निवडणे समाविष्ट आहे. हे खर्च कमी करू शकते आणि सॅम्पलिंगची कार्यक्षमता वाढवू शकते.

क्लस्टर सॅम्पलिंगचे प्रकार:
एकदाच सॅम्पलिंग (One-stage cluster sampling): एक किंवा अधिक गट निवडले जातात आणि त्या गटांमधून सर्व सदस्य निवडले जातात.
द्विस्तरीय सॅम्पलिंग (Two-stage cluster sampling): प्रथम गट निवडले जातात, आणि नंतर त्या गटांमधून काही सदस्य निवडले जातात.
Python मध्ये क्लस्टर सॅम्पलिंग कसा करायचा?
pandas आणि random मॉड्यूलचा वापर करून आपण क्लस्टर सॅम्पलिंग सहज करू शकतो.
1. क्लस्टर सॅम्पलिंग (One-Stage)
उदाहरणार्थ, समजा आपल्याकडे एका शाळेतील १०० वर्गांच्या (clusters) लिस्ट आहे, आणि आपल्याला त्या वर्गांमधून काही निवडक वर्ग निवडायचे आहेत.
कोड:

```
python
Copy code
import pandas as pd
import random
# डेटा तयार करा (वर्गाचे उदाहरण)
clusters = {
   'Class': ['Class_A', 'Class_B', 'Class_C', 'Class_D', 'Class_E', 'Class_F', 'Class_G', 'Class_H',
'Class_I', 'Class_J'],
   'Students': [30, 28, 35, 40, 50, 60, 45, 30, 25, 35]  # प्रत्येक वर्गातील विद्यार्थ्यांची संख्या
}
# DataFrame तयार करा
df = pd.DataFrame(clusters)
# गटांची निवड (क्लस्टर सॅम्पलिंग: 3 गट निवडा)
selected_clusters = random.sample(df['Class'].tolist(), 3)
# निवडक गटांची माहिती
print("selected_clusters:", selected_clusters)
# त्या गटातील सर्व विद्यार्थ्यांची माहिती:
selected_data = df[df['Class'].isin(selected_clusters)]
print("\nAll students Infromation in selected_clusters:\n", selected_data)
```

सारांश:
random.sample() पद्धतीचा वापर करून, १० गटांमधून ३ गट निवडले जातात.
नंतर, त्या निवडक गटांमधील सर्व विद्यार्थ्यांची माहिती isin() फंक्शनचा वापर करून घेतली जाते.
आउटपुट:

less

Copy code

निवडक गट: ['Class_A', 'Class_C', 'Class_F']
निवडक गटांमधील सर्व विद्यार्थ्यांची माहिती:

	Class	Students
0	Class_A	30
2	Class_C	35
5	Class_F	60

2. क्लस्टर सॅम्पलिंग (Two-Stage)

दुसऱ्या उदाहरणात, आपण द्विस्तरीय सॅम्पलिंग करू. त्यात, प्रथम गट (उदाहरणार्थ, वर्ग) निवडले जातील आणि त्यात काही विद्यार्थ्यांची निवड केली जाईल.
कोड:

python

Copy code

```
import pandas as pd
import random
# डेटा तयार करा (वर्ग व विद्यार्थ्यांचा उदाहरण)
data = {
    'Class': ['Class_A', 'Class_A', 'Class_A', 'Class_B', 'Class_B', 'Class_B',
              'Class_C', 'Class_C', 'Class_C', 'Class_D', 'Class_D', 'Class_D'],
    'Student_ID': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12],
    'Name': ['John', 'Alice', 'Bob', 'Eve', 'Charlie', 'David',
             'Grace', 'Harry', 'Ivy', 'Jack', 'Kathy', 'Leo']
}
# DataFrame तयार करा
df = pd.DataFrame(data)
# 2-स्टेज क्लस्टर सॅम्पलिंग:
# 1. गटांची निवड (3 गट निवडा)
selected_classes = random.sample(df['Class'].unique().tolist(), 3)
print("Selected Class:", selected_classes)
# 2. निवडक गटांमधून काही विद्यार्थ्यांची निवड
selected_students = df[df['Class'].isin(selected_classes)]
# प्रत्येक निवडक गटामधून 2 विद्यार्थी निवडा
final_sample = selected_students.groupby('Class').apply(lambda x: x.sample(2))
```

print("\nSelected Students information in Two Stage Sampling:\n", final_sample)
सारांश:
प्रथम स्टेज: random.sample() वापरून ३ गट निवडले जातात.
दुसरी स्टेज: प्रत्येक निवडक गटामधून २ विद्यार्थ्यांची निवड groupby().apply().sample() पद्धतीने केली जाते.
आउटपुट:
less
Copy code
निवडक गट: ['Class_A', 'Class_C', 'Class_B']
द्विस्तरीय सॅम्पलिंगनंतर निवडक विद्यार्थ्यांची माहिती:

Class	Student	ID	Name
Class_A	0	1	John
Class_A	1	2	Alice
Class_C	7	8	Harry
Class_C	8	9	Ivy
Class_B	3	4	Eve
Class_B	4	5	Charlie

निष्कर्ष:
क्लस्टर सॅम्पलिंग मध्ये, आपल्याला विविध गटांमधून काही गट निवडायचे असतात.
द्विस्तरीय सॅम्पलिंग मध्ये, एक पाऊल पुढे जाऊन प्रत्येक गटामधून काही सदस्य निवडले जातात.
याचा उपयोग साधारणतः मोठ्या लोकसंख्येवर किंवा पसरलेल्या लोकसंख्येवर केल्यास कार्यक्षम होतो.
केसेस वापरा(Use Cases): क्लस्टर सॅम्पलिंगचा उपयोग क्षेत्रीय प्रयोगांमध्ये किंवा सर्वेक्षणांमध्ये(surveys) केला जातो जेथे लोकसंख्या भौगोलिकदृष्ट्या विखुरलेली असते.

5. Over(ओव्हर) sampling and Under (उंडर) sampling: ओव्हरसॅम्पलिंग आणि अंडरसॅम्पलिंग ही डेटासेटमधील वर्ग असमतोल दूर करण्यासाठी वापरली जाणारी तंत्रे आहेत. ओव्हरसॅम्पलिंगमध्ये अल्पसंख्याक वर्गातील उदाहरणांची संख्या वाढवणे समाविष्ट असते, तर अंडरसॅम्पलिंगमध्ये बहुसंख्य वर्गातील उदाहरणांची संख्या कमी करणे समाविष्ट असते.
ओव्हर सॅम्पलिंग (Over-sampling): ज्या वर्गाच्या उदाहरणांची संख्या कमी आहे, त्याच्या उदाहरणांची संख्या वाढवली जाते.
अंडर सॅम्पलिंग (Under-sampling): ज्या वर्गाच्या उदाहरणांची संख्या जास्त आहे, त्याची संख्या कमी केली जाते.
आता, Python मध्ये ओव्हर सॅम्पलिंग आणि अंडर सॅम्पलिंग कसे करायचे ते पाहू.

1. ओव्हर सॅम्पलिंग (Over-sampling)

ओव्हर सॅम्पलिंग मध्ये कमी असलेल्या वर्गाच्या उदाहरणांची संख्या वाढवली जाते, उदाहरणार्थ, वर्ग A मध्ये १०० रेकॉर्ड्स आहेत, तर वर्ग B मध्ये १० रेकॉर्ड्स आहेत, तर वर्ग B मध्ये आणखी रेकॉर्ड्स जोडले जातात.
Python कोड (SMOTE वापरून ओव्हर सॅम्पलिंग):
Python मध्ये imblearn लायब्ररीत SMOTE (Synthetic Minority Over-sampling Technique) वापरून ओव्हर सॅम्पलिंग करता येते.
python

Copy code

```
# आवश्यक लायब्ररीज आयात करा
import pandas as pd
from imblearn.over_sampling import SMOTE
from sklearn.datasets import make_classification
# उदाहरणार्थ डेटासेट तयार करा
X, y = make_classification(n_samples=1000, n_features=20, n_classes=2, weights=[0.1, 0.9],
flip_y=0, random_state=42)
# डेटासेटची श्रेणी आणि आकार
print("Over Sampling:", pd.Series(y).value_counts())
# SMOTE वापरून ओव्हर सॅम्पलिंग करा
smote = SMOTE(random_state=42)
X_res, y_res = smote.fit_resample(X, y)
# पुनः सॅम्पलिंग नंतर वर्ग वितरण
print("Over Sampling:", pd.Series(y_res).value_counts())
```

सारांश:

make_classification(): एक नमुना डेटासेट तयार करण्यासाठी वापरले जाते, जिथे एका वर्गाची संख्या कमी असते.

SMOTE(): कमी असलेल्या वर्गाच्या उदाहरणांची संख्या वाढवते आणि सिम्युलेटेड (कृत्रिम) उदाहरण तयार करते.

fit_resample(): हे फंक्शन डेटा पुनः सॅम्पल करते.

आउटपुट:

go

Copy code

पूर्व सॅम्पलिंग वर्ग वितरण:

1	900
0	100

dtype: int64

पुन्हा सॅम्पलिंग वर्ग वितरण:

1 900

0 900

dtype: int64

2. अंडर सॅम्पलिंग (Under-sampling)

अंडर सॅम्पलिंग मध्ये जास्त असलेल्या वर्गाच्या उदाहरणांची संख्या कमी केली जाते, उदाहरणार्थ, वर्ग A मध्ये १०० रेकॉर्ड्स आहेत, तर वर्ग B मध्ये १००० रेकॉर्ड्स आहेत, तर वर्ग B मधील काही रेकॉर्ड्स काढले जातात.

Python कोड (Random Under-sampling वापरून अंडर सॅम्पलिंग):

python

Copy code

```
# आवश्यक लायब्ररीज आयात करा
import pandas as pd
from imblearn.under_sampling import RandomUnderSampler
from sklearn.datasets import make_classification
# उदाहरणार्थ डेटासेट तयार करा
X, y = make_classification(n_samples=1000, n_features=20, n_classes=2,
                    weights=[0.9, 0.1], flip_y=0, random_state=42)
# डेटासेटची श्रेणी आणि आकार
print("Under Sampling:", pd.Series(y).value_counts())
# RandomUnderSampler वापरून अंडर सॅम्पलिंग करा
under_sampler = RandomUnderSampler(random_state=42)
X_res, y_res = under_sampler.fit_resample(X, y)
# पुनः सॅम्पलिंग नंतर वर्ग वितरण
print("Under Sampling:", pd.Series(y_res).value_counts())
```

सारांश:

RandomUnderSampler(): जास्त असलेल्या वर्गाच्या उदाहरणांची संख्या कमी करते.

fit_resample(): या पद्धतीने डेटा पुनः सॅम्पल केला जातो.

आउटपुट:

```
go
Copy code
पूर्व सॅम्पलिंग वर्ग वितरण:
1    900
0    100
dtype: int64
पुन्हा सॅम्पलिंग वर्ग वितरण:
1    100
0    100
dtype: int64
```

ओव्हर सॅम्पलिंग आणि अंडर सॅम्पलिंग मध्ये फरक:

ओव्हर सॅम्पलिंग:

कमी असलेल्या वर्गाची संख्या वाढवली जाते.

साधारणतः SMOTE किंवा RandomOverSampler वापरून करण्यात येते.

यामुळे सॅम्पलिंगमधील विविधता (diversity) सुधारली जाते, कारण काही वेळा SMOTE कृत्रिम डेटा उदाहरण तयार करतो.

अंडर सॅम्पलिंग:

जास्त असलेल्या वर्गाची संख्या कमी केली जाते.

साधारणतः RandomUnderSampler वापरून करण्यात येते.

यामुळे मोठ्या वर्गाच्या डेटाची प्रमाणिकता (representativeness) कमी होऊ शकते, कारण डेटा कमी केला जातो.

निष्कर्ष:

ओव्हर सॅम्पलिंग आणि अंडर सॅम्पलिंग दोन्हीच असमान वर्ग वितरण (imbalanced data) सुधारण्यासाठी उपयोगी पद्धती आहेत.

ओव्हर सॅम्पलिंग कमी असलेल्या वर्गाचा डेटा वाढवतो, तर अंडर सॅम्पलिंग जास्त असलेल्या वर्गाचा डेटा कमी करतो.

SMOTE आणि RandomOverSampler/RandomUnderSampler हे दोन प्रमुख तंत्र आहेत जे पायथनमध्ये ओव्हर आणि अंडर सॅम्पलिंगसाठी वापरले जातात.

केसेस वापरा(Use Cases): या पद्धती वर्गीकरणाच्या समस्यांमध्ये महत्त्वपूर्ण आहेत जेथे एक वर्ग लक्षणीयरीत्या दुसऱ्या वर्गापेक्षा जास्त आहे, जसे की क्रेडिट कार्ड फसवणूक शोधणे.

5.1.6 दर्जात्मक त्रुटी (Standard Error/ स्टँडर्ड एरर): स्टँडर्ड एरर हे परिवर्तनशीलतेचा अंदाज घेण्यासाठी सांख्यिकीमध्ये वापरल्या जाणाऱ्या गणिती साधनांपैकी एक आहे. त्याचे संक्षिप्त रूप SE असे आहे. आकडेवारीची प्रमाणित त्रुटी किंवा पॅरामीटरचा अंदाज म्हणजे त्याच्या नमुना वितरणाचे मानक विचलन. त्या मानक विचलनाचा अंदाज म्हणून आपण त्याची व्याख्या करू शकतो.

दर्जात्मक त्रुटी सूत्र (Standard Error Formula): दर्जात्मक त्रुटी सूत्र

लोकसंख्येचे वर्णन करणाऱ्या नमुन्याची अचूकता SE सूत्राद्वारे ओळखली जाते. नमुन्याचा अर्थ जो दिलेल्या लोकसंख्येपासून विचलित होतो आणि ते विचलन असे दिले जाते;

$$SE_{\bar{x}} = \frac{S}{\sqrt{n}}$$

जेथे S हे मानक विचलन (standard deviation) आहे आणि n ही निरीक्षणांची संख्या (number of observations)आहे.

अंदाजाची मानक त्रुटी (Standard Error of Estimate) (SEE):अंदाजाची प्रमाणित त्रुटी म्हणजे कोणत्याही अंदाजाच्या अचूकतेचा अंदाज. हे SEE म्हणून दर्शविले जाते. प्रतिगमन रेषा अंदाजाच्या वर्ग विचलनाच्या बेरजेचे अवमूल्यन करते. याला स्क्वेअर एररची बेरीज म्हणून देखील ओळखले जाते. SEE हे सरासरी वर्ग विचलनाचे वर्गमूळ आहे. अपेक्षित मूल्यांमधून काही अंदाजांचे विचलन अंदाज सूत्राच्या मानक त्रुटीद्वारे दिले जाते.

$$\boldsymbol{SEE = \sqrt{\frac{(x_i - \bar{x})^2}{n-2}}}$$

5.1) दिलेल्या डेटाच्या मानक त्रुटीची (standard error) गणना करा:

y: 5, 10, 12, 15, 20

ऊत्तराची: प्रथम आपल्याला दिलेल्या डेटाचा मध्य शोधायचा आहे;

सरासरी(mean) = (5+10+12+15+20)/5 = 62/5 = 10.5

आता, मानक विचलन म्हणून गणना केली जाऊ शकते;

S = दिलेल्या डेटाचे प्रत्येक मूल्य आणि सरासरी मूल्य/मूल्यांची संख्या यांच्यातील फरकाची बेरीज.

त्यामुळे,

$$S = \sqrt{\frac{(5-10.5)^2 + (10-10.5)^2 + (12-10.5)^2 + (15-10.5)^2 + (20-10.5)^2}{5}}$$

S = 5.35

म्हणून, SE चा अंदाज सूत्राद्वारे केला जाऊ शकतो;

$$SE = \frac{S}{\sqrt{n}} = \frac{5.35}{\sqrt{5}} = 2.39$$

5.2) नमुना डेटासाठी मानक त्रुटी शोधा: 2, 3, 4, 5, 6.

ऊत्तराची: प्रथम आपल्याला दिलेल्या डेटाचा मध्य शोधायचा आहे;

सरासरी(mean) = $\frac{2+3+4+5+6}{5} = \frac{20}{5} = 4$

आता, मानक विचलन म्हणून गणना केली जाऊ शकते;

S = दिलेल्या डेटाचे प्रत्येक मूल्य आणि सरासरी मूल्य/मूल्यांची संख्या यांच्यातील फरकाची बेरीज.

त्यामुळे,

$$S = \sqrt{\frac{(2-4)^2 + (3-4)^2 + (4-4)^2 + (5-4)^2 + (6-4)^2}{5}}$$

$S = 1.265$

म्हणून, SE चा अंदाज सूत्राद्वारे केला जाऊ शकतो;

$$SE = \frac{S}{\sqrt{n}} = \frac{1.265}{\sqrt{5}} = 0.566$$

5.3) उदाहरण - मानक त्रुटीचे गणित

समजा, एक शाळेतील 100 विद्यार्थ्यांची वयाची माहिती घेतली आहे. वयाच्या लोकसंख्येचा मानक विचलन σ=5 आहे. जर आपण 25 विद्यार्थ्यांचा नमुना घेतला आणि त्या नमुन्याचा आकार n=25 असेल, तर मानक त्रुटी किती असेल?

गणना:

सुरुवातीला, सूत्र वापरून मानक त्रुटीची गणना करूया:

$$SE = \frac{S}{\sqrt{n}} = \frac{5}{\sqrt{25}} = \frac{5}{5} = 1$$

उत्तर: मानक त्रुटी SE=1 असेल.

5.4) उदाहरण - मानक त्रुटीचे गणित दुसऱ्या पद्धतीने

समजा, एक कंपनी उत्पादन करत आहे आणि त्यांच्या उत्पादनांची वजनाची माहिती घेतली आहे. त्या उत्पादनाच्या वजनाचा मानक विचलन σ=2 किलो आहे. जर आपण १०० उत्पादनांचा नमुना घेतला असेल, तर मानक त्रुटी किती असेल?

गणना:

सुरुवातीला, सूत्र वापरून मानक त्रुटीची गणना करूया:

$$SE = \frac{S}{\sqrt{n}} = \frac{2}{\sqrt{100}} = \frac{2}{10} = 0.2$$

उत्तर: मानक त्रुटी SE=0.2 किलो असेल.

5.5) उदाहरण - मानक त्रुटी आणि नमुन्याचा आकार

समजा, एका शहरात ५०० लोकांची ऊंची मोजली आहे आणि त्यांचा मानक विचलन σ=8सेंटीमीटर आहे. आपल्याला ५० लोकांचा नमुना घ्यायचा आहे. नमुन्याचा आकार n=50 असेल, तर मानक त्रुटी किती असेल?

गणना:

$$SE = \frac{S}{\sqrt{n}} = \frac{8}{\sqrt{50}} = \frac{8}{7.017} \approx \mathbf{1.13}$$

उत्तर: मानक त्रुटी SE≈1.13 असेल.

4. मानक त्रुटीचा अर्थ

मानक त्रुटी मोजण्याचा उद्देश म्हणजे नमुन्यांच्या सरासरींचे वितरण आणि लोकसंख्येतील प्रत्यक्ष सरासरी यामध्ये असलेला फरक. अधिक नमुने घेतल्यास, मानक त्रुटी कमी होईल कारण आपण एक अधिक विश्वसनीय अंदाज तयार करत आहोत.

उदाहरणार्थ:

मानक त्रुटी कमी असणे म्हणजे आपल्याकडे जास्त अचूक माहिती आहे.

मानक त्रुटी जास्त असणे म्हणजे आपल्या अंदाजांमध्ये जास्त चुकांची शक्यता आहे.

5. मानक त्रुटी आणि नमुना आकाराचा संबंध

मानक त्रुटी आणि नमुन्याचा आकार यामध्ये एक महत्त्वाचा संबंध आहे:

नमुन्याचा आकार जास्त असेल, तर मानक त्रुटी कमी होईल.

नमुन्याचा आकार कमी असेल, तर मानक त्रुटी जास्त होईल.

उदाहरण:

समजा, आपल्या असमान वर्गीकरणाच्या डेटामध्ये २० लोकांची माहिती घेऊन मानक त्रुटी काढली, आणि त्यानंतर 100 लोकांची माहिती घेतली. अधिक लोकसंख्येचा नमुना घेणे म्हणजे आपल्याला अधिक अचूक परिणाम मिळतील.

गणना:

20 लोकांसाठी मानक त्रुटी:

$$SE_1 = \frac{S}{\sqrt{n}} = \frac{5}{\sqrt{20}} = \frac{5}{4.47} \approx \mathbf{1.12}$$

100 लोकांसाठी मानक त्रुटी:

$$SE_2 = \frac{S}{\sqrt{n}} = \frac{5}{\sqrt{100}} = \frac{5}{10} = \mathbf{0.5}$$

इथे आपण पाहू शकता की, नमुन्याचा आकार वाढविल्यामुळे मानक त्रुटी कमी झाली आहे.

निष्कर्ष:

मानक त्रुटी (SE) हा एक महत्त्वाचा संख्याशास्त्रीय घटक आहे, जो लोकसंख्येच्या सरासरीसाठी एक अंदाज देतो.

मानक त्रुटीच्या मापाने आपल्याला कळते की, आपला नमुना किती अचूक आहे.

अधिक नमुने घेतल्यास मानक त्रुटी कमी होईल, आणि अधिक अचूक परिणाम मिळवता येतील.

5.2 हायपोथिसिस(Hypothesis): हायपोथिसिस चाचणीचा वापर नमुना डेटा वापरून गृहीतकेच्या संभाव्यतेचे मूल्यांकन करण्यासाठी केला जातो. चाचणी डेटा दिल्यास, गृहीतकाच्या तर्कशुद्धतेबद्दल पुरावे प्रदान करते. सांख्यिकीय विश्लेषक विश्लेषित केल्या जात असलेल्या लोकसंख्येच्या याद्दच्छिक नमुन्याचे मोजमाप करून आणि परीक्षण करून गृहीतकांची चाचणी घेतात.

5.2.1 नल हायपोथिसिस (H_0) / Null Hypothesis: नल हायपोथिसिस (H0) - याचा विचार गर्भित गृहीतक म्हणून केला जाऊ शकतो. "शून्य" म्हणजे "काहीच नाही." हे गृहितक सांगते की गटांमध्ये फरक नाही किंवा व्हेरिएबल्समध्ये कोणताही संबंध नाही. शून्य गृहीतक हे यथास्थितीचे गृहितक आहे किंवा कोणताही बदल नाही.

5.2.2 :अल्टरनेटीव हायपोथिसिस (H_a) / Alternative Hypothesis - अल्टरनेटीव हायपोथिसिस याला दावा असेही म्हणतात. या गृहीतकाने या विषयावरील तुमच्या संशोधनाच्या आधारे, डेटा दाखवण्याची तुमची अपेक्षा आहे हे सांगितले पाहिजे. आपण असेही म्हणू शकतो की तो फक्त शून्याचा पर्याय आहे. गृहीतक चाचणीमध्ये, पर्यायी सिद्धांत हे विधान आहे ज्याची संशोधक चाचणी करत आहे. हे विधान संशोधकाच्या दृष्टिकोनातून खरे आहे आणि शेवटी ते पर्यायी गृहीतकाने बदलण्यासाठी नल नाकारणे सिद्ध होते. या गृहीतकामध्ये, दोन किंवा अधिक व्हेरिएबल्समधील फरक संशोधकांनी वर्तवला आहे, जसे की चाचणीमध्ये आढळलेल्या डेटाचा नमुना संधीमुळे नाही.

उदाहरणे:

नल हायपोथिसिस ($\mathbf{H_0}$): कारखाना कामगारांच्या पगारात लिंगाच्या आधारावर कोणताही फरक नाही.

अल्टरनेटीव हायपोथिसिस ($\mathbf{H_a}$): हा पुरुष कारखाना कामगारांना महिला कारखान्यातील कामगारांपेक्षा जास्त पगार असतो.

नल हायपोथिसिस ($\mathbf{H_0}$): उंची आणि बुटाच्या आकारात कोणताही संबंध नाही.

अल्टरनेटीव हायपोथिसिस ($\mathbf{H_a}$): उंची आणि बुटाच्या आकारामध्ये सकारात्मक संबंध आहे.

नल हायपोथिसिस ($\mathbf{H_0}$): कामावरील अनुभवाचा वीट दगडी बांधकामाच्या गुणवत्तेवर कोणताही परिणाम होत नाही.

अल्टरनेटीव हायपोथिसिस ($\mathbf{H_a}$): वीट गवंडीच्या कामाच्या गुणवत्तेवर नोकरीच्या अनुभवाचा प्रभाव पडतो.

नल आणि अल्टरनेटीव हायपोथिसिस मधील फरक (Difference Between Null and Alternative Hypothesis)

नल हायपोथिसिस	अल्टरनेटीव हायपोथिसिस
हे सूचित करते की दोन मोजलेल्या घटनांमध्ये कोणताही संबंध नाही.	हे एक गृहितक आहे की यादृच्छिक(random) कारणामुळे निरीक्षण केलेल्या डेटावर किंवा नमुन्यावर परिणाम होऊ शकतो.
हे $\mathbf{H_0}$ द्वारे दर्शविले जाते	हे $\mathbf{H}_a$ किंवा $\mathbf{H_1}$ द्वारे दर्शविले जाते
उदाहरण: रोहन लकी ड्रॉमध्ये किमान रु.100000 जिंकेल	उदाहरण: रोहन लकी ड्रॉमध्ये रु.100000 पेक्षा कमी जिंकेल.

मुळात, अल्टरनेटीव हायपोथिसिस चे तीन प्रकार आहेत, ते आहेत;

डावी- टेलेड (**Left-Tailed**): येथे, नमुना प्रमाण (π) निर्दिष्ट मूल्यापेक्षा कमी असणे अपेक्षित आहे जे π_0 ने दर्शविले जाते, जसे की;

$\mathbf{H_1} : \pi < \pi_0$

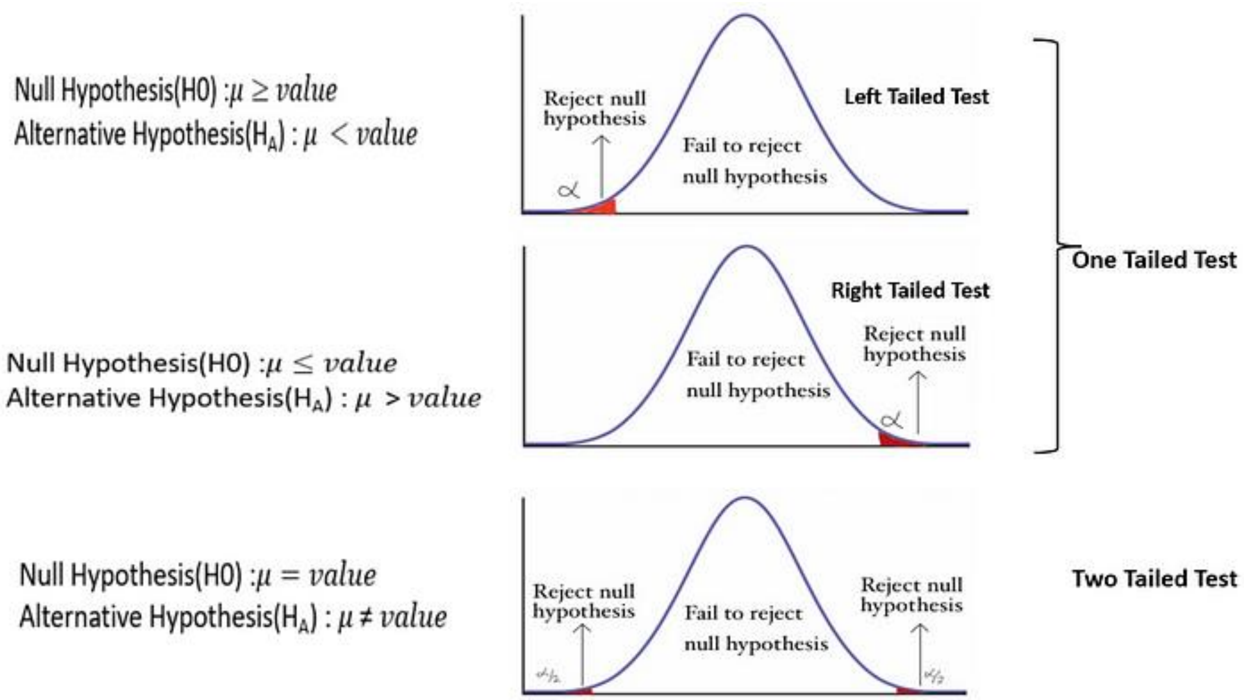
उजवे- टेलेड (**Right-Tailed**): हे दर्शविते की नमुना प्रमाण (π) काही मूल्यापेक्षा मोठे आहे, π_0 ने दर्शविले जाते.

$\mathbf{H_1} : \pi > \pi_0$

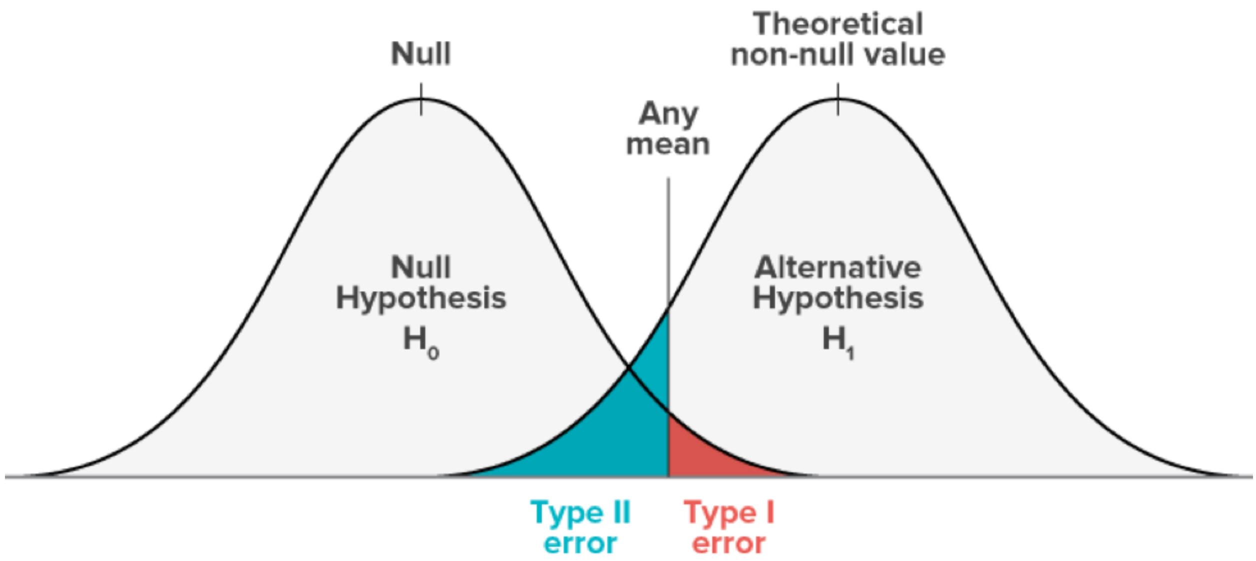
दोन- टेलेड (**Two-Tailed**): या गृहीतकानुसार, नमुना प्रमाण (π द्वारे दर्शविलेले) विशिष्ट मूल्याच्या बरोबरीचे नाही जे π_0 ने दर्शवले जाते.

$\mathbf{H_1} : \pi \neq \pi_0$

टीप: तीनही पर्यायी गृहितकांसाठी शून्य गृहितक, $\mathbf{H_1} : \pi = \boldsymbol{\pi_0}$ असेल.



Type I ('α' म्हणून देखील ओळखले जाते) आणि Type II ('β' म्हणून देखील ओळखले जाते) त्रुटी (Errors):



एक प्रकार I त्रुटी (खोटे-सकारात्मक) जर एखाद्या अन्वेषकाने लोकसंख्येमध्ये खरे असलेली शून्य गृहितक नाकारली तर उद्भवते; एक प्रकार II त्रुटी (खोटे-नकारात्मक) उद्भवते जर अन्वेषक एक शून्य गृहितक नाकारण्यात अयशस्वी झाले जे प्रत्यक्षात लोकसंख्येमध्ये खोटे आहे.

5.3 लेव्हल ऑफ सिग्निफिकन्से (महत्त्वाची पातळी /Level of Significance):

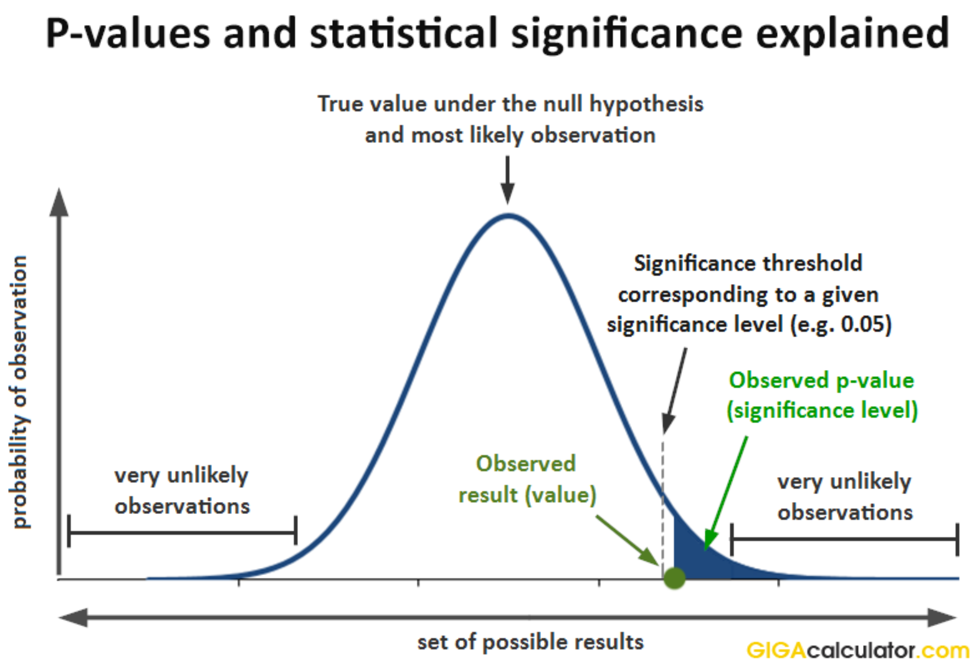
महत्त्वाच्या पातळीची व्याख्या शून्य गृहीतके(नल हायपोथिसिस) चुकीच्या निर्मूलनाची निश्चित संभाव्यता म्हणून केली जाते जेव्हा ते खरे असते. महत्त्वाची पातळी म्हणजे सांख्यिकीय महत्त्व मोजणे. शून्य गृहीतक स्वीकारले जाईल की नाकारले जाईल हे ते परिभाषित करते. शून्य गृहितक खोटे किंवा नाकारले जाण्यासाठी परिणाम सांख्यिकीयदृष्ट्या महत्त्वपूर्ण आहे की नाही हे ओळखणे अपेक्षित आहे.

महत्त्वाची पातळी ग्रीक चिन्ह α (अल्फा) द्वारे दर्शविली जाते. म्हणून, महत्त्वाची पातळी खालीलप्रमाणे परिभाषित केली आहे:

पी-मूल्य(p-value) काय आहे?

सांख्यिकीय गृहीतक चाचणीमध्ये, p-मूल्य (संभाव्यता मूल्य Probability Value) हे दिलेल्या सांख्यिकीय चाचणीचे शून्य गृहितक सत्य असताना निरीक्षण केलेले किंवा अधिक टोकाचे परिणाम शोधण्याचे संभाव्यतेचे माप आहे. p-मूल्य हे एक प्राथमिक मूल्य आहे जे परिकल्पना चाचणीच्या परिणामांचे सांख्यिकीय महत्त्व मोजण्यासाठी वापरले जाते.

लेव्हल ऑफ सिग्निफिकन्से = p (प्रकार I त्रुटी) = α



लेव्हल ऑफ सिग्निफिकन्स

0.05 किंवा 5% घेतली जाते. जेव्हा p-मूल्य(p-value) कमी असते, तेव्हा त्याचा अर्थ असा होतो की मान्यताप्राप्त मूल्ये सुरुवातीला गृहीत धरलेल्या लोकसंख्येच्या मूल्यापेक्षा लक्षणीय भिन्न आहेत. p-मूल्य (p-value) शक्य तितके कमी असल्यास ते अधिक महत्त्वपूर्ण असल्याचे म्हटले जाते. तसेच, p-मूल्य(p-value) खूपच कमी असल्यास परिणाम अत्यंत महत्त्वपूर्ण असेल. परंतु, सामान्यतः, ०.०५ पेक्षा लहान p-मूल्ये(p-value) महत्त्वाची म्हणून ओळखली जातात, कारण p-मूल्य ०.०५ पेक्षा कमी मिळणे फारच कमी सराव आहे.

निकालाच्या सांख्यिकीय महत्त्वाची पातळी मोजण्यासाठी, अन्वेषकाने प्रथम p-मूल्याची गणना करणे आवश्यक आहे. हे एक प्रभाव ओळखण्याची संभाव्यता परिभाषित करते जे प्रदान करते की शून्य गृहितक सत्य आहे.

जेव्हा p-मूल्य महत्त्वाच्या पातळीपेक्षा कमी असते (α), शून्य गृहीतक(नल हायपोथिसिस) नाकारले जाते. असे निरीक्षण केलेले p-मूल्य α पेक्षा कमी नसल्यास, सैद्धांतिकदृष्ट्या शून्य गृहितक(नल हायपोथिसिस) स्वीकारले जाते. परंतु व्यावहारिकदृष्ट्या, आम्ही अनेकदा नमुन्याच्या आकाराचा आकार वाढवतो आणि आम्ही महत्त्वाच्या पातळीवर पोहोचतो की नाही ते तपासतो. 10% च्या महत्त्वाच्या स्तरावर आधारित p-मूल्याची
सामान्य व्याख्या:

जर $p > 0.1$ असेल, तर शून्य गृहीतकासाठी(नल हायपोथिसिस) कोणतेही गृहितक(assumption) असणार नाही

$p > 0.05$ आणि $p \le 0.1$ असल्यास, याचा अर्थ शून्य गृहीतकांसाठी कमी गृहीतक असेल.

जर $p > 0.01$ आणि $p \le 0.05$ असेल, तर शून्य गृहीतकाबद्दल एक मजबूत गृहितक असणे आवश्यक आहे.

जर $p \le 0.01$ असेल, तर शून्य गृहीतकाबद्दल एक अतिशय मजबूत गृहितक सूचित केले आहे.

5.3.1 टेस्ट ऑफ सिग्निफिकन्स (Test of Significance / महत्त्वाची चाचणी): महत्त्वाची चाचणी ही निरीक्षण केलेल्या डेटाची दाव्याशी तुलना करण्याची औपचारिक प्रक्रिया आहे (ज्याला गृहीतक देखील म्हटले जाते), ज्याच्या सत्याचे मूल्यांकन केले जात आहे. दावा हे पॅरामीटरबद्दलचे विधान आहे, जसे की लोकसंख्येचे प्रमाण p किंवा लोकसंख्येचा अर्थ μ.

एकदा निरीक्षणात्मक अभ्यास किंवा प्रयोगाद्वारे नमुना डेटा गोळा केला गेला की, सांख्यिकीय निष्कर्ष विश्लेषकांना पुराव्याचे मूल्यांकन करण्यास किंवा ज्या लोकसंख्येवरून नमुना काढला गेला आहे त्याबद्दलच्या काही दाव्यांचे मूल्यांकन करण्यास अनुमती देते. नमुना डेटावर आधारित दाव्यांना समर्थन देण्यासाठी किंवा नाकारण्यासाठी वापरल्या जाणाऱ्या अनुमानांच्या पद्धती महत्त्वाच्या चाचण्या म्हणून ओळखल्या जातात.

महत्त्वाची प्रत्येक चाचणी शून्य गृहितक $\mathbf{H_0}$ ने सुरू होते. $\mathbf{H_0}$ हा एक सिद्धांत दर्शवितो जो पुढे मांडला गेला आहे, एकतर तो सत्य आहे असे मानले जाते किंवा कारण ते युक्तिवादाचा आधार म्हणून वापरले जाणार आहे, परंतु ते सिद्ध झालेले नाही. उदाहरणार्थ, नवीन औषधाच्या क्लिनिकल चाचणीमध्ये, शून्य गृहीतक असू शकते की नवीन औषध सध्याच्या औषधापेक्षा सरासरी चांगले नाही. आम्ही $\mathbf{H_0}$ लिहू: दोन औषधांमध्ये सरासरी फरक नाही.

सांख्यिकीय महत्त्वासाठी सूत्र (Formula for statistical significance)

$$\boldsymbol{Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}}$$

जेथे X̄ हा नमुना सरासरी(sample mean) आहे, μ हा लोकसंख्येचा मध्य(population mean) आहे, σ हा प्रमाणित विचलन आहे आणि n हा नमुना आकार(sample size) आहे

5.3.2 Confidence limit: आत्मविश्वास मर्यादा म्हणजे आत्मविश्वास मध्यांतराच्या वरच्या आणि खालच्या टोकावरील संख्या

उदाहरणार्थ, जर तुमची सरासरी 7.4 असेल 5.4 च्या आत्मविश्वास मर्यादेसह आणि ९.४, तुमचा आत्मविश्वास मध्यांतर 5.4 आहेते 9.4

. बहुतेक लोक 95% वापरतात

आत्मविश्वास मर्यादा, जरी तुम्ही इतर मूल्ये वापरू शकता. ९५% सेट करत आहे

आत्मविश्वास मर्यादा म्हणजे जर तुम्ही लोकसंख्येमधून वारंवार यादृच्छिक नमुने घेतले आणि प्रत्येक नमुन्यासाठी सरासरी आणि आत्मविश्वास मर्यादांची गणना केली तर, 95% साठी आत्मविश्वास मध्यांतर

तुम्ही भरपूर प्रमाणात काम करत असल्यास, वेगवेगळ्या नमुन्याच्या आकारांसाठी आत्मविश्वास मर्यादांची अंदाजे कल्पना असणे चांगले आहे, त्यामुळे तुम्हाला विशिष्ट तुलनासाठी किती डेटा आवश्यक आहे याची कल्पना आहे.

50% जवळच्या प्रमाणांसाठी, आत्मविश्वास मध्यांतर अंदाजे ±30%,10%,3% आणि 1% n=10,100,1000 आणि 10,000 साठी अनुक्रमे आहेत. म्हणूनच राजकीय पोलमध्ये "एररचे मार्जिन", ज्याचा नमुना साधारणतः 1,000 च्या आसपास असतो, साधारणतः 3% असतो. अर्थात, ही ढोबळ कल्पना प्रत्यक्ष शक्ती विश्लेषणाला पर्याय नाही.

n	proportion=0.10	proportion=0.50
10	0.0025, 0.4450	0.1871, 0.8129
100	0.0490, 0.1762	0.3983, 0.6017
1000	0.0821, 0.1203	0.4685, 0.5315
10,000	0.0942, 0.1060	0.4902, 0.5098

Z चाचणी (Z Test):

Z चाचणी ही एक सांख्यिकीय चाचणी आहे जी अंदाजे सामान्य वितरणाचे अनुसरण करणाऱ्या डेटावर केली जाते. z चाचणी एका नमुन्यावर, दोन नमुन्यांवर किंवा परिकल्पना चाचणीसाठी प्रमाणानुसार केली जाऊ शकते. जेव्हा लोकसंख्येतील फरक ओळखला जातो तेव्हा दोन मोठ्या नमुन्यांचे माध्यम वेगळे आहेत की नाही हे ते तपासते.

डेटाच्या पॅरामीटर्सच्या आधारावर z चाचणीचे पुढे डावी-पुच्छ, उजवे-पुच्छ आणि दोन-पुच्छ गृहीतक चाचण्यांमध्ये वर्गीकरण केले जाऊ शकते. या लेखात, आपण z चाचणी, त्याचे सूत्र, z चाचणी आकडेवारी आणि उदाहरणे वापरून विविध प्रकारच्या डेटासाठी चाचणी कशी करावी याबद्दल अधिक जाणून घेऊ.

5.4Test of Significance for Large Sample (मोठ्या नमुन्यासाठी महत्त्वाची चाचणी): Z चाचणी सामान्यतः मोठ्या नमुन्यांसाठी लागू केली जाते. झेड-चाचणी ही साधारणपणे वितरित केलेल्या डेटासाठी सांख्यिकीय चाचणी आहे. केंद्रीय मर्यादा प्रमेयामुळे, अनेक चाचणी आकडेवारी साधारणपणे मोठ्या नमुन्यांसाठी वितरीत केली जाते. लोकसंख्येतील भिन्नता (मानक विचलन) ज्ञात आहे.

हायपॉथीसिस टेस्टिंगचे सामान्य टप्पे:

हायपॉथीसिस सेट करा (H0 आणि H1).

योग्य सांख्यिकीय चाचणी निवडा (उदाहरणार्थ, t-चाचणी, z-चाचणी).

नमुन्याचा डेटा संकलित करा.

p-मूल्य (p-value) काढा.

निर्णय घ्या:

p-मूल्य α (सामान्यतः 0.05) पेक्षा कमी असल्यास H0 नाकारून H1 स्वीकारा.

p-मूल्य α पेक्षा जास्त असल्यास H0 नाकारू नका.

साधी t-चाचणी (One Sample t-Test):

5.6)समजा, एका कंपनीने दावा केला आहे की त्यांच्या उत्पादनाचे वजन ५०० ग्रॅम आहे. आपण ३० उत्पादनांची चाचणी घेतो आणि त्या उत्पादनांची सरासरी वजन ४९५ ग्रॅम आहे. आपण पद्धतशीरपणे चाचणी करू इच्छिता की त्यांचा दावा खरा आहे की नाही.

दिलेले:

नल हायपॉथीसिस (H0): μ=500 (वजन ५०० ग्रॅम आहे)

पर्यायी हायपॉथीसिस (H1): μ≠500 (वजन ५०० ग्रॅम नाही)

नमुन्याचा आकार (n) = 30

सरासरी वजन ($\bar{x}$) = 495 ग्रॅम

लोकसंख्येचा मानक विचलन (s) = 10 ग्रॅम

महत्वाची पातळी (α) = 0.05

t-चाचणी सूत्र:

$$\boldsymbol{t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}}$$

गणना:

$$\mathrm{t} = \frac{495 - 500}{\frac{10}{\sqrt{30}}} \approx -2.47$$

निर्णय: तुम्हाला t-तक्त्याच्या वापरातून p-मूल्य शोधावे लागेल. t=−2.74 आणि df=29 (स्वतंत्रता डिग्री n−1) वापरून, p-मूल्य साधारणतः 0.01 येईल.

तुम्ही सेट केलेली महत्वाची पातळी (α) 0.05 आहे. कारण p-मूल्य 0.01 α पेक्षा कमी आहे, त्यामुळे H0 नाकारला जाईल आणि H1 स्वीकारला जाईल. याचा अर्थ, कंपनीचा दावा खोटा आहे आणि उत्पादनाचे वजन 500 ग्रॅम नाही.

5.7) एका कारखान्याचा दावा आहे की पिठाच्या पिशवीचे सरासरी वजन 500 ग्रॅम आहे. 50 पिशव्यांचा याद्दच्छिक नमुना घेतला जातो आणि नमुना सरासरी वजन 495 ग्रॅम असल्याचे आढळले. लोकसंख्या मानक विचलन 15 ग्रॅम म्हणून ओळखले जाते. नमुना सरासरी 0.05 महत्त्वाच्या पातळीवर दावा केलेल्या 500 ग्रॅमच्या सरासरीपेक्षा लक्षणीय भिन्न आहे का ते तपासा.

नल हायपॉथीसिस (H0): μ=500 (वजन ५०० ग्रॅम आहे)

पर्यायी हायपॉथीसिस (H1): μ≠500 (वजन ५०० ग्रॅम नाही)

नमुन्याचा आकार (n) = 50

सरासरी वजन ($\bar{x}$) = 495 ग्रॅम

लोकसंख्येचा मानक विचलन (s) = 15 ग्रॅम

महत्वाची पातळी (α) = 0.05

t-चाचणी सूत्र:

$$\boldsymbol{t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}}$$

$$\boldsymbol{t = \frac{495 - 500}{\frac{15}{\sqrt{50}}} \cong -2.36}$$

येथे दोन-पुच्छ चाचणीसाठी महत्त्वपूर्ण मूल्य शोधा

α=0.05

±1.96, हे सह द्वि-पुच्छ चाचणीसाठी Z-वितरण सारणीमधील महत्त्वपूर्ण मूल्य

α =0.05 आहे

|Z|=2.36 हे 1.96 पेक्षा मोठे आहे, आम्ही शून्य परिकल्पना नाकारतो.

निष्कर्ष:

0.05 महत्त्वाच्या पातळीवर हे निष्कर्ष काढण्यासाठी पुरेसे पुरावे आहेत की नमुना सरासरी 500 ग्रॅमपेक्षा लक्षणीय भिन्न आहे.

दोन स्वतंत्र नमुन्यांची t-चाचणी (Two-sample t-test):

5.8) समजा, एका शाळेतील दोन वर्गांच्या विद्यार्थ्यांची सरासरी गुणांची तुलना केली जात आहे. वर्ग A आणि वर्ग B यांच्या सरासरी गुणांमध्ये फरक आहे की नाही हे तपासण्यासाठी हायपॉथीसिस टेस्टिंग करा.

दिलेले:

नल हायपॉथीसिस (H0): μ1=μ2 (वर्ग A आणि B मध्ये सरासरी गुण समान आहेत)

पर्यायी हायपॉथीसिस (H1): μ1≠μ2 (वर्ग A आणि B मध्ये सरासरी गुण वेगळे आहेत)

वर्ग A चा नमुना आकार $n_1 = 40$, वर्ग B चा नमुना आकार n_2=50

वर्ग A ची सरासरी $\overline{x_1}$=75, वर्ग B ची सरासरी $\overline{x_2}$=70

वर्ग A चा मानक विचलन σ_1=12, वर्ग B चा मानक विचलन σ_2=15

महत्वाची पातळी (α) = 0.05

दोन नमुन्यांची t-चाचणी सूत्र:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{(\sigma_1)^2}{(n_1)^2} + \frac{(\sigma_2)^2}{(n_2)^2}}}$$

गणना:

$$t = \frac{(75-70)}{\sqrt{\frac{(12)^2}{(40)^2}+\frac{(15)^2}{(50)^2}}} = t = \frac{5}{\sqrt{\frac{144}{1600}+\frac{225}{2500}}} \approx 1.76$$

निर्णय:

आपण t-तक्त्याचा वापरून p-मूल्य शोधतो. t=1.76 आणि df (स्वतंत्रता डिग्री) काढण्यासाठी:

df=min (n_1−1, n_2−1)=min (40−1,50−1)=39

तुम्ही t=1.76 आणि df=39 असताना p-मूल्य साधारणतः 0.085 येईल.

तुम्ही सेट केलेली महत्वाची पातळी (α) 0.05 आहे. कारण p-मूल्य 0.085 α पेक्षा जास्त आहे, त्यामुळे H0 नाकारला जाणार नाही, म्हणजे वर्ग A आणि वर्ग B मध्ये सरासरी गुण समान आहेत.

Z-चाचणी (Z-test):

5.9) समजा, एका शहरातील लोकसंख्येची उंची १६० सेंटीमीटर आहे. एक संस्थानाच्या विद्यार्थी समूहाच्या उंचीचा सर्वेक्षण घेतला आहे आणि ते १५५ सेंटीमीटर आहे. आपल्याला हायपॉथीसिस टेस्टिंगद्वारे तपासायचं आहे की विद्यार्थ्यांचा उंचीचा औसत लोकसंख्येच्या सरासरीपेक्षा वेगळा आहे की नाही.

दिलेले:

नल हायपॉथीसिस (H0): μ=160 (विद्यार्थ्यांचा औसत उंची १६० सेंटीमीटर आहे)

पर्यायी हायपॉथीसिस (H1): μ≠160 (विद्यार्थ्यांचा औसत उंची १६० सेंटीमीटर नाही)

नमुन्याचा आकार (n) = 50
सरासरी उंची $\bar{x}$= 155 सेंटीमीटर
लोकसंख्येचा मानक विचलन σ = 12 सेंटीमीटर
महत्वाची पातळी (α) = 0.05
Z-चाचणी सूत्र:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

गणना:

$$z = \frac{155 - 160}{\frac{12}{\sqrt{50}}} = \frac{-5}{\frac{12}{7.071}} \approx -2.94$$

निर्णय:
Z-चाचणी प्रमाणे, Z=−2.94. Z-तक्त्याच्या वापराने p-मूल्य मिळवता येईल. ZZ=−2.94 असल्याने, p-मूल्य 0.003 येईल.
पॉझिटिव्ह p-मूल्य 0.003 आहे, जे α = 0.05 पेक्षा कमी आहे, त्यामुळे H0H_0H0 नाकारला जातो. म्हणजे, विद्यार्थ्यांचा उंची लोकसंख्येच्या सरासरीपेक्षा वेगळा आहे.
निष्कर्ष:
हायपॉथीसिस टेस्टिंग हा एक सांख्यिकी प्रक्रिया आहे ज्यात नल हायपॉथीसिस H0H_0H0 आणि पर्यायी हायपॉथीसिस H1H_1H1 चा तपास केला जातो.
t-चाचणी, z-चाचणी, आणि F-चाचणी अशा विविध प्रकारांच्या चाचण्यांचा वापर करून परीक्षण केले जाते.
p-मूल्य आधारित निर्णय घेतला जातो, ज्या आधारावर हायपॉथीसिस स्वीकारली किंवा नाकारली जाते.

5.5 Sampling distribution of proportion (प्रमाणाचे नमुना वितरण): नमुना प्रमाणांचे संकलन संभाव्यता वितरण तयार करते ज्याला नमुना प्रमाणाचे नमुना वितरण म्हणतात. नमुना प्रमाणांच्या वितरणाचा मध्य, $\mu_{\hat{p}}$ दर्शविला जातो, लोकसंख्येच्या प्रमाणात असतो.

Suppose random samples of size n are drawn from a population in which the proportion with a characteristic of interest is p. The mean $\mu_{\hat{p}}$ and standard deviation $\sigma_{\hat{p}}$ of the sample proportion $\hat{p}$ satisfy

$\mu_{\hat{p}} = p$ and $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$

where q=1−p.
The Sampling Distribution of the Sample Proportion
For large samples, the sample proportion is approximately normally distributed, with mean $\mu_{\hat{p}} = p$ and standard deviation $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$
A sample is large if the interval [p−3$\sigma_{\hat{p}}$, p+3$\sigma_{\hat{p}}$] lies wholly within the interval [0,1].

5.6Comparison of large samples (मोठ्या नमुन्यांची तुलना)

मोठ्या नमुन्यांची तुलना करताना (N>३०), आम्ही विशेषत: दोन किंवा अधिक गटांच्या साधनांची किंवा प्रमाणांची तुलना करण्यावर लक्ष केंद्रित करतो आणि त्यांच्यामध्ये लक्षणीय फरक आहेत का हे पाहण्यासाठी. तुलनेसाठी वापरल्या जाणाऱ्या पद्धती अनेकदा Z-चाचण्या आणि t-चाचण्यांवर अवलंबून असतात, जरी मोठ्या नमुन्यांसाठी, Z-चाचणीला सामान्यतः सेंट्रल लिमिट प्रमेय (CLT) मुळे प्राधान्य दिले जाते जे असे सांगते की नमुना वितरणाचा

अर्थ (किंवा प्रमाण)) जेव्हा नमुना आकार मोठा असतो तेव्हा अंदाजे सामान्य होते.खाली मोठ्या नमुन्यांची तुलना करण्यासाठी सर्वात सामान्य चाचण्यांची तुलना आहे

झेड टेस्ट म्हणजे काय?

z चाचणी ही एक चाचणी आहे जी दोन लोकसंख्येची साधने भिन्न आहेत की नाही हे तपासण्यासाठी वापरली जाते की डेटा सामान्य वितरणानुसार प्रदान केला जात नाही. या उद्देशासाठी, शून्य गृहितक आणि पर्यायी गृहितक सेट करणे आवश्यक आहे आणि z चाचणी आकडेवारीचे मूल्य मोजले जाणे आवश्यक आहे. निर्णय निकष z गंभीर मूल्यावर आधारित आहे.

Z चाचणी (झेड टेस्ट) फॉर्म्युला

दोन लोकसंख्येच्या साधनांमध्ये फरक आहे की नाही हे तपासण्यासाठी z चाचणी सूत्र z गंभीर मूल्याशी z आकडेवारीची तुलना करते. गृहीतक चाचणीमध्ये, z क्रिटिकल व्हॅल्यू वितरण आलेखाला स्वीकृती आणि नकार क्षेत्रांमध्ये विभाजित करते. जर चाचणी आकडेवारी नाकारण्याच्या प्रदेशात आली तर शून्य गृहितक नाकारले जाऊ शकते अन्यथा ते नाकारले जाऊ शकत नाही. एक नमुना आणि दोन-नमुना z चाचणीसाठी आवश्यक परिकल्पना चाचण्या सेट करण्यासाठी z चाचणी सूत्र खाली दिले आहे.

एक-नमुना Z चाचणी(One-Sample Z Test)

लोकसंख्येचे मानक विचलन ज्ञात असताना नमुना सरासरी आणि लोकसंख्येच्या मध्यामध्ये फरक आहे का हे तपासण्यासाठी एक-नमुना z चाचणी वापरली जाते. z चाचणी आकडेवारीचे सूत्र खालीलप्रमाणे दिले आहे:

$$Z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

$\bar{x}$ is the sample mean (नमुना सरासरी)

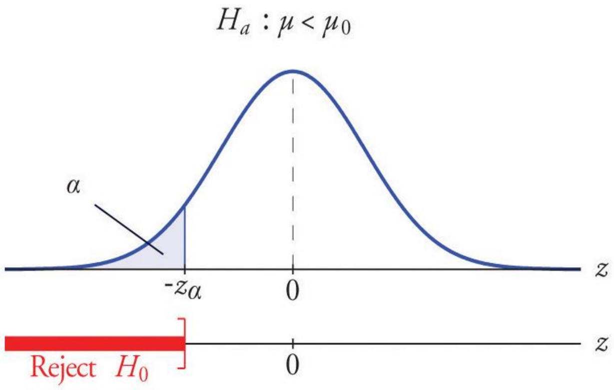
μ is the population mean (लोकसंख्येच्या मध्या)

s is the population standard deviation and

n is the sample size.

z चाचणी आकडेवारीवर आधारित एक नमुना z चाचणी सेट करण्यासाठी अल्गोरिदम खालीलप्रमाणे दिलेला आहे:

डावी- टेलेड चाचणी (**Left-Tailed Test**):



Null Hypothesis: $H_0 : \mu = \mu_0$

Alternate Hypothesis: $H_1 : \mu < \mu_0$

Decision Criteria: If the z statistic < z critical value then reject the null hypothesis.

लेफ्ट टेल्ड टेस्ट मध्ये, आपला परिभाषा असा असतो:

नल हायपॉथीसिस (H0): μ≥μ0 (म्हणजे, नमुन्याचा पॅरामिटर, सामान्यत: सरासरी, ठराविक मूल्याच्या किंवा त्यापेक्षा जास्त आहे).

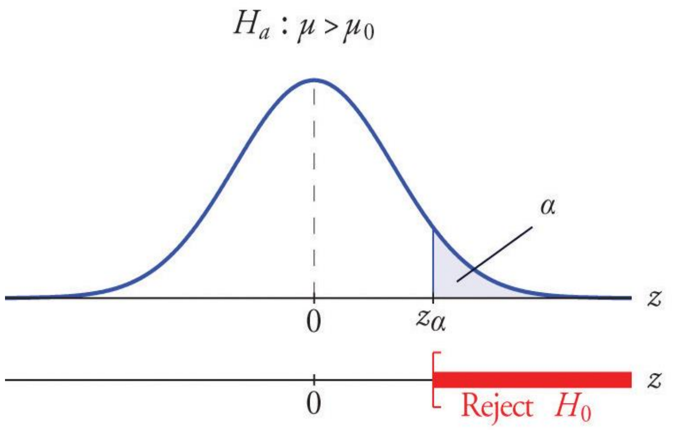
पर्यायी हायपॉथीसिस (H1): $\mu<\mu0$ (म्हणजे, नमुन्याचा पॅरामिटर ठराविक मूल्यापेक्षा कमी आहे).
उदाहरण:
समजा, एका कार निर्मात्या कंपनीचा दावा आहे की त्यांचा कारचा गॅसोलीन कंझम्पशन (mileage) १५ किमी/लिटर आहे. आपल्याला तपासायचं आहे की, त्या कारचा गॅसोलीन कंझम्पशन १५ किमी/लिटरपेक्षा कमी आहे की नाही.
दिलेले:
नल हायपॉथीसिस (H0): $\mu=15$(कारचा गॅसोलीन कंझम्पशन १५ किमी/लिटर आहे)
पर्यायी हायपॉथीसिस (H1): $\mu<15$ (कारचा गॅसोलीन कंझम्पशन १५ किमी/लिटरपेक्षा कमी आहे)
नमुन्याचा आकार (n) = 25
नमुन्याची सरासरी ($\boldsymbol{x}$) = 14.4 किमी/लिटर
लोकसंख्येचा मानक विचलन (s) = 1.2 किमी/लिटर
महत्वाची पातळी (α) = 0.05
पद्धत:
1. t-चाचणीचे सूत्र:
लेफ्ट टेल्ड टेस्टमध्ये, आपल्याला t-चाचणी वापरायची आहे. सूत्र आहे: $t = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}$
जिथे:
$\bar{x}$ = नमुन्याची सरासरी,
μ_0 = नल हायपॉथीसिस असलेली ठराविक मूल्य,
s = लोकसंख्येचा मानक विचलन,
n = नमुन्याचा आकार.
2. गणना:
आता, आपल्याला दिलेल्या माहितीचा वापर करून ttt-तक्त्याचा वापर करण्यासाठी t-चाचणीची गणना करू.

$$t = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = t = \frac{14.4 - 15}{\frac{1.2}{\sqrt{25}}} = -2.5$$

3. t-मूल्य आणि p-मूल्य:
आपल्याला $t=-2.5$ आणि $n-1=24$ या स्वतंत्रता डिग्रीसाठी t-तक्ता वापरून p-मूल्य शोधावे लागेल. $t=-2.5$आणि df = 24 असताना, p-मूल्य साधारणतः 0.01 असू शकते.
4. निर्णय:
आपण ५% महत्वाची पातळी (α = 0.05) ठेवली आहे. पद्धतीप्रमाणे:
जर p-मूल्य α पेक्षा कमी असेल (अर्थात $p<0.05$) तर नल हायपॉथीसिस (H0) नाकारली जाईल आणि पर्यायी हायपॉथीसिस (H1) स्वीकारली जाईल.
जर p-मूल्य α पेक्षा जास्त असेल, तर H0 स्वीकारली जाईल.
आता, पद्धतीनुसार, आपल्याला p-मूल्य = 0.01 दिले गेले आहे, जे $\alpha=0.05$ पेक्षा कमी आहे. त्यामुळे नल हायपॉथीसिस नाकारली जाते आणि पर्यायी हायपॉथीसिस स्वीकारली जाते. याचा अर्थ, आपल्याला हे सिद्ध झाले आहे की, कारचा गॅसोलीन कंझम्पशन १५ किमी/लिटरपेक्षा कमी आहे.
निष्कर्ष:
नल हायपॉथीसिस: $\mu=15$ (कारचा गॅसोलीन कंझम्पशन १५ किमी/लिटर आहे)

पर्यायी हायपॉथीसिस: μ<15 (कारचा गॅसोलीन कंझम्पशन १५ किमी/लिटरपेक्षा कमी आहे)
t-मूल्य: -2.5
p-मूल्य: 0.01
निर्णय: p-मूल्य α=0.05 पेक्षा कमी असल्यामुळे नल हायपॉथीसिस नाकारली जाते आणि पर्यायी हायपॉथीसिस स्वीकारली जाते, म्हणजे कारचा गॅसोलीन कंझम्पशन १५ किमी/लिटरपेक्षा कमी आहे.

Right Tailed Test उजवे- टेलेड चाचणी:

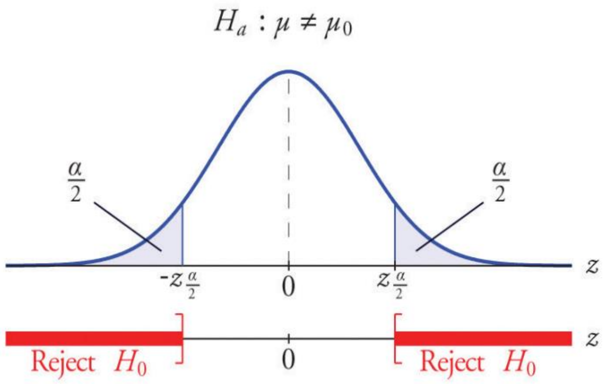


Null Hypothesis: $H_0: \mu = \mu_0$

Alternate Hypothesis: $H_1: \mu > \mu_0$

Decision Criteria: If the z statistic > z critical value then reject the null hypothesis.

Two Tailed Test दोन- टेलेड चाचणी**:**



Null Hypothesis: $H_0: \mu = \mu_0$

Alternate Hypothesis: $H_1: \mu \neq \mu_0$

Decision Criteria: If the z statistic > z critical value then reject the null hypothesis.
राइट-टेल्ड टेस्ट चे परिभाषा:
नल हायपॉथीसिस (H0): μ≤μ0 (पॅरामिटर एका ठराविक मूल्यापेक्षा जास्त नाही)
पर्यायी हायपॉथीसिस (H1): μ>μ0 (पॅरामिटर एका ठराविक मूल्यापेक्षा जास्त आहे)
5.10)समजा, एका कंपनीने दावा केला आहे की, त्यांच्या उत्पादकतेचा दर (productivity rate) प्रति दिवस 80 युनिट आहे. आपल्याला तपासायचं आहे की, त्यांची उत्पादकता 80 युनिट्स पेक्षा जास्त आहे का.
दिलेले:
नल हायपॉथीसिस (H0): μ=80 (उत्पादकतेचा दर 80 युनिट्स आहे)
पर्यायी हायपॉथीसिस (H1): μ>80 (उत्पादकतेचा दर 80 युनिट्स पेक्षा जास्त आहे)
नमुन्याचा आकार n=30
नमुन्याची सरासरी $\bar{x}$ =82 युनिट्स

लोकसंख्येचा मानक विचलन σ=5 युनिट्स
महत्वाची पातळी (α) = 0.05
पद्धत:
1. Z-चाचणीचे सूत्र:
राइट-टेल्ड टेस्टसाठी, आपण Z-चाचणी वापरणार आहोत. Z-चाचणीचे सूत्र असे आहे:
जिथे:
$\bar{x}$= नमुन्याची सरासरी,
μ0 = नल हायपॉथीसिस असलेली ठराविक मूल्य,
σ = लोकसंख्येचा मानक विचलन,
n = नमुन्याचा आकार.
2. गणना:
आता, आपल्याला दिलेल्या माहितीचा वापर करून Z-चाचणीची गणना करू.

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = t = \frac{82 - 80}{\frac{5}{\sqrt{30}}} \approx 2.19$$

3. Z-मूल्य आणि p-मूल्य:
आता, आपल्याला Z = 2.19 मिळाले आहे. Z-तक्त्याच्या वापराने p-मूल्य शोधू शकता.
Z=2.19 च्या p-मूल्य किमान 0.014 असू शकते.
4. निर्णय:
आपण महत्वाची पातळी (α=0.05) सेट केली आहे.
जर p-मूल्य α पेक्षा कमी असेल (म्हणजे p<0.05) तर नल हायपॉथीसिस (H0) नाकारली जाते आणि पर्यायी हायपॉथीसिस (H1) स्वीकारली जाते.
जर p-मूल्य α पेक्षा जास्त असेल, तर H0 स्वीकारली जाईल.
आता, पद्धतीनुसार, आपल्याला p-मूल्य = 0.014 मिळालं आहे, जे α = 0.05 पेक्षा कमी आहे. त्यामुळे नल हायपॉथीसिस नाकारली जाते आणि पर्यायी हायपॉथीसिस स्वीकारली जाते.
निष्कर्ष:
नल हायपॉथीसिस (H0): μ=80 (उत्पादकतेचा दर ८० युनिट्स आहे)
पर्यायी हायपॉथीसिस (H1): μ>80(उत्पादकतेचा दर ८० युनिट्स पेक्षा जास्त आहे)
Z-मूल्य: 2.19
p-मूल्य: 0.014
निर्णय: p-मूल्य α=0.05 पेक्षा कमी आहे, त्यामुळे नल हायपॉथीसिस नाकारली जाते आणि पर्यायी हायपॉथीसिस स्वीकारली जाते, म्हणजे कंपनीचा उत्पादन दर ८० युनिट्स पेक्षा जास्त आहे.
दोन नमुना Z चाचणी
दोन सॅम्पलच्या माध्यमांमध्ये फरक आहे का हे तपासण्यासाठी दोन नमुना z चाचणी वापरली जाते. z चाचणी सांख्यिकी सूत्र खालीलप्रमाणे दिले आहे:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(\sigma_1)^2}{(n_1)^2} + \frac{(\sigma_2)^2}{(n_2)^2}}}$$

$\bar{x}_1, \mu_1, (\sigma_1)^2$ are the sample mean, population mean and population variance respectively for the first sample. $\bar{x}_2, \mu_2, (\sigma_2)^2$ are the sample mean, population mean and population variance respectively for the second sample.

उदाहरण:

समजा, दोन वेगवेगळ्या स्कूल्स (School A आणि School B) च्या विद्यार्थ्यांची गणित (Math) विषयातील सरासरी गुण (average marks) तपासायची आहेत. आपल्याला हे तपासायचं आहे की दोन्ही स्कूल्सच्या विद्यार्थ्यांची गणितातील सरासरी गुण वेगवेगळी आहेत का.

दिलेले:

School A:

सरासरी ($\bar{x}_1$) = 78

लोकसंख्येचा मानक विचलन (σ_1) = 10

नमुन्याचा आकार (n_1) = 50

School B:

सरासरी ($\overline{x_2}$) = 75

लोकसंख्येचा मानक विचलन (σ_2) = 12

नमुन्याचा आकार (n_2) = 60

नल हायपॉथीसिस (H0): μ1=μ2 (दोन्ही स्कूल्सच्या विद्यार्थ्यांची सरासरी गुण समान आहेत)

महत्वाची पातळी (α) = 0.05

पद्धत:

Z-चाचणीचे सूत्र वापरा: $Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(\sigma_1)^2}{(n_1)^2} + \frac{(\sigma_2)^2}{(n_2)^2}}}$

नल हायपॉथीसिस नुसार, $\mu_1 = \mu_2$, त्यामुळे $(\mu_1 - \mu_2) = 0$

$$Z = \frac{(78 - 75) - 0}{\sqrt{\frac{(10)^2}{(50)^2} + \frac{(12)^2}{(60)^2}}} \approx 1.43$$

3. p-मूल्य शोधा:

Z = 1.43 आहे.

Z-तक्ता वापरून p-मूल्य शोधा. Z = 1.43 साठी p-मूल्य साधारणतः 0.0762 आहे.

4. निर्णय:

जर p-मूल्य α=0.05 पेक्षा जास्त असेल, तर नल हायपॉथीसिस स्वीकारली जाईल.

जर p-मूल्य α=0.05 पेक्षा कमी असेल, तर नल हायपॉथीसिस नाकारली जाईल.

इथे p-मूल्य = 0.0762 आहे, जे α=0.05 पेक्षा जास्त आहे. त्यामुळे नल हायपॉथीसिस स्वीकारली जाते

5.7 विद्यार्थ्यांचे टी वितरण (Student's t distribution)

टी चाचणी करण्यासाठी, तुम्ही नमुन्यासाठी टीची गणना करा आणि त्याची तुलना गंभीर मूल्याशी करा. योग्य क्रिटिकल व्हॅल्यू शोधण्यासाठी, तुम्हाला स्वातंत्र्याच्या योग्य डिग्रीसह विद्यार्थ्यांचे टी वितरण वापरण्याची आवश्यकता आहे.

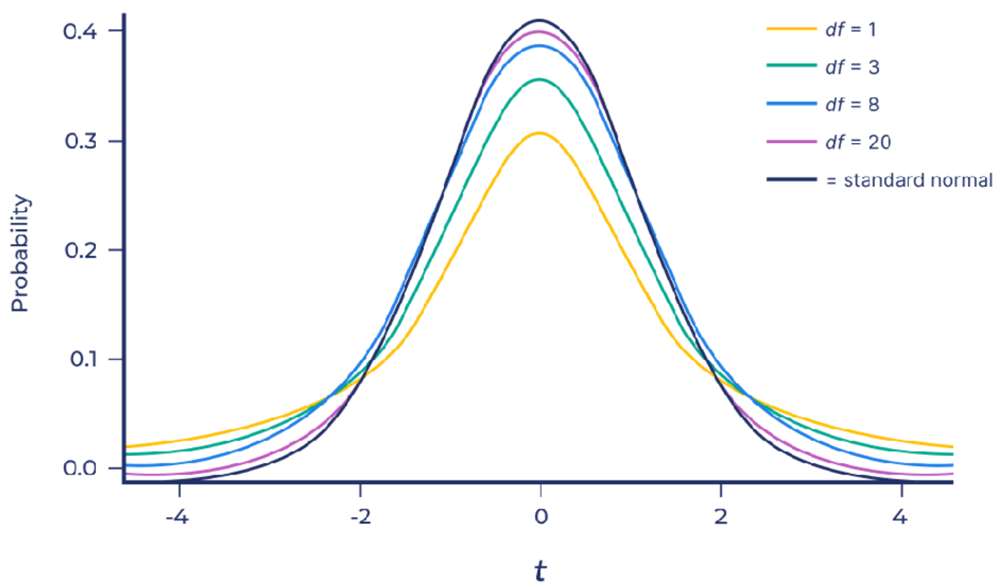
विद्यार्थ्यांच्या टीचे शून्य वितरण स्वातंत्र्याच्या अंशांसह बदलते:

जेव्हा df = 1, वितरण जोरदार लेप्टोकर्टिक असते, म्हणजे अत्यंत मूल्यांची संभाव्यता सामान्य वितरणापेक्षा जास्त असते.

जसजसे डीएफ वाढते तसतसे वितरण अरुंद आणि कमी लेप्टोकर्टिक होते. हे मानक सामान्य वितरणासारखेच वाढते

जेव्हा df ≥ 30, विद्यार्थ्यांचे t वितरण साधारण सामान्य वितरणासारखेच असते.

जर तुमच्याकडे नमुना आकार 30 पेक्षा जास्त असेल, तर तुम्ही विद्यार्थ्यांच्या t वितरणाऐवजी मानक सामान्य वितरण (ज्याला z वितरण म्हणूनही ओळखले जाते) वापरू शकता.



सांख्यिकीच्या स्वातंत्र्याचे अंश(degrees of freedom in statistics) म्हणजे नमुना आकार वजा निर्बंधांची संख्या. बऱ्याच वेळा, निर्बंध हे पॅरामीटर्स असतात ज्यांचा अंदाज सांख्यिकी मोजण्यासाठी मध्यवर्ती चरण म्हणून केला जातो.

D_f =n − r

Where

n नमुना आकार (sample size) आहे

r ही निर्बंधांची संख्या आहे, सामान्यत: अंदाजित पॅरामीटर्सच्या संख्येइतकीच

स्वातंत्र्याचे प्रमाण नकारात्मक असू शकत नाही. परिणामी, तुम्ही अंदाजित केलेल्या पॅरामीटर्सची संख्या तुमच्या नमुना आकारापेक्षा मोठी असू शकत नाही.

Test-specific formulas:

Test	Formula	Notes
One-sample t test	$df = n - 1$	
Independent samples t test	$df = n1 + n2 - 2$	Where $n1$ is the sample size of group 1 and $n2$ is the sample size of group 2

Test	Formula	Notes
Dependent samples *t* test	$df = n - 1$	Where *n* is the number of pairs
Simple linear regression	$df = n - 2$	
Chi-square goodness of fit test	$df = k - 1$	Where *k* is the number of groups
Chi-square test of independence	$df = (r - 1) * (c - 1)$	Where *r* is the number of rows (groups of one variable) and *c* is the number of columns (groups of the other variable) in the contingency table
One-way ANOVA	Between-group $df = k - 1$ Within-group $df = N - k$ Total $df = N - 1$	Where *k* is the number of groups and *N* is the sum of all groups' sample sizes

विद्यार्थ्यांचा गुणासंबंधी T-चाचणी

5.11) एका शाळेतील 25 विद्यार्थ्यांचा सरासरी गुण 75% आहेत आणि त्याची मानक विचलन 10% आहे. शाळेतील नवीन पद्धतीचा उपयोग करून विद्यार्थ्यांची सरासरी गुण तपासण्याची आवश्यकता आहे. आपल्याला हे तपासायचं आहे की नवीन पद्धतीचा उपयोग केलेल्या विद्यार्थ्यांचा गुण 75% च्या सरासरीपेक्षा जास्त आहे का?

चरण 1: H_0 आणि H_1 ठरवा

H_0: नवीन पद्धतीचा उपयोग करणाऱ्या विद्यार्थ्यांचा गुण 75% आहे.

H_1: नवीन पद्धतीचा उपयोग करणाऱ्या विद्यार्थ्यांचा गुण 75% पेक्षा जास्त आहे.

चरण 2: सूत्र आणि माहिती

न = 25 (विद्यार्थ्यांची संख्या)

सरासरी गुण (x̄) = 80%

सरासरी गुण (μ) = 75%

मानक विचलन (s) = 10%

T-चाचणीचे सूत्र:

$t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$

जिथे:

$\bar{x}$= नमुना सरासरी (Sample mean)

μ = जनसंख्या सरासरी (Population mean)

s = नमुना मानक विचलन (Sample standard deviation)

n = नमुना आकार (Sample size)

चरण ३: गणना करा

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{80 - 75}{\frac{10}{\sqrt{25}}} = 2.5$$

चरण 4: T-Distribution किंव्हा टी-टेबल वापरा

डिग्री ऑफ फ्रीडम (df) = n - 1 = 25 - 1 = 24

1% पातळीवर T-टेबलमध्ये 24 डिग्री ऑफ फ्रीडमसाठी t मूल्य सुमारे 2.492 आहे.
चरण 5: निर्णय घ्या
t मूल्य (2.5) हे T-टेबलमध्ये दिलेल्या t मूल्यापेक्षा जास्त आहे (2.492).
त्यामुळे H_0 नाकारता येईल आणि निष्कर्ष काढता येईल की नवीन पद्धतीने विद्यार्थ्यांचा गुण 75% पेक्षा जास्त आहे.
कडक तपासणीची चाचणी
5.12)धरणा करा की एका कंपनीला 10 किव्हा 15 नोकऱ्यांसाठी मानक वेतनाचे प्रमाण तपासायचं आहे. त्यांनी 10 कर्मचाऱ्यांवर आधारित एक नमुना घेतला आहे. या 10 कर्मचाऱ्यांच्या वेतनाचे विश्लेषण करा. नंतर 5% पातळीवर T-चाचणी वापरून ह्याचा निर्णय घ्या.
माहिती:
नमुना आकार (n) = 10
नमुना सरासरी वेतन ($\bar{x}$) = ₹ 20000
जनसंख्या सरासरी वेतन (μ) = ₹ 18000
मानक विचलन (s) = ₹ 2500
चरण 1: H_0 आणि H_1 ठरवा
H_0: नमुना सरासरी वेतन जनसंख्येच्या सरासरीच्या समान आहे.
H_1: नमुना सरासरी वेतन जनसंख्येच्या सरासरीपेक्षा वेगळं आहे.
चरण 2: T-चाचणी सूत्र वापरून गणना करा

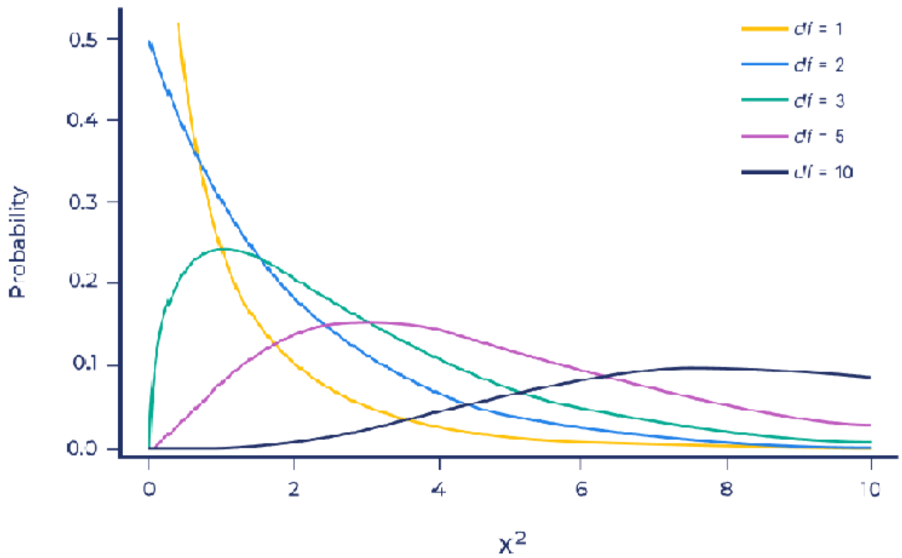
$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{20000 - 18000}{\frac{2500}{\sqrt{10}}} = 2.53$$

चरण 3: T-टेबल वापरा
डिग्री ऑफ फ्रीडम (df) = n - 1 = 10 - 1 = 9
5% पातळीवर T-टेबलमध्ये ९ डिग्री ऑफ फ्रीडमसाठी t मूल्य सुमारे 2.262 आहे.
चरण 4: निर्णय घ्या
t मूल्य (2.53) हे T-टेबलमधील t मूल्यापेक्षा जास्त आहे (2.262).
त्यामुळे H_0 नाकारता येईल आणि निष्कर्ष काढता येईल की नमुना सरासरी वेतन जनसंख्येच्या सरासरीपेक्षा वेगळं आहे.
प्रोडक्ट गुणवत्तेची चाचणी
5.13) समजा की एका फॅक्टरीने 150 उत्पादनांच्या गुणवत्तेचा नमुना घेतला आहे. 150 उत्पादनांमध्ये 100 उत्पादनांचे वजन 100 ग्रॅम आहे, आणि 50 उत्पादनांचे वजन 105 ग्रॅम आहे. आपल्याला 5% पातळीवर तपासायचं आहे की उत्पादनांचे वजन 100 ग्रॅम आहे का?
माहिती:
नमुना आकार (n) = 150
नमुना सरासरी ($\bar{x}$) = 102 ग्रॅम
जनसंख्या सरासरी (μ) = 100 ग्रॅम
मानक विचलन (s) = 2 ग्रॅम
चरण 1: H_0 आणि H_1 ठरवा
H_0: उत्पादनांचे वजन 100 ग्रॅम आहे.

H_1: उत्पादनांचे वजन 100 ग्रॅम नाही.
चरण 2: T-चाचणी सूत्र वापरून गणना करा
$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{102-100}{\frac{2}{\sqrt{150}}}$=12.27
चरण 3: T-टेबल वापरा
डिग्री ऑफ फ्रीडम (df) = n - 1 = 150 - 1 = 149
5% पातळीवर T-टेबलमध्ये १४९ डिग्री ऑफ फ्रीडमसाठी t मूल्य अंदाजे 1.976 आहे.
चरण 4: निर्णय घ्या
t मूल्य (12.27) हे T-टेबलमधील t मूल्यापेक्षा जास्त आहे (1.976).
त्यामुळे H_0 नाकारता येईल आणि निष्कर्ष काढता येईल की उत्पादनांचे वजन 100 ग्रॅम नाही.

5.8 ची-चौरस वितरण(Chi-square distribution): ची-स्क्वेअर चाचणी करण्यासाठी, तुम्ही नमुन्याच्या ची-स्क्वेअरची तुलना महत्त्वपूर्ण मूल्याशी करता. योग्य क्रिटिकल व्हॅल्यू शोधण्यासाठी, तुम्हाला ची-स्क्वेअर डिस्ट्रिब्युशनचा वापर स्वातंत्र्याच्या योग्य अंशांसह करणे आवश्यक आहे.

ची-स्क्वेअरचे शून्य वितरण स्वातंत्र्याच्या अंशांसह बदलते, परंतु विद्यार्थ्यांच्या टी वितरणापेक्षा वेगळ्या प्रकारे:
जेव्हा df < 3, संभाव्यता वितरणाचा आकार मागे "J" सारखा असतो.
जेव्हा df ≥ 3, संभाव्यता वितरण कुबडाच्या आकाराचे असते, कुबड्याचे शिखर X2 = df – 2 वर असते. कुबडा उजवा-तिरका असतो, याचा अर्थ वितरण त्याच्या शिखराच्या उजव्या बाजूला लांब असते.
जेव्हा df > 90, ची-स्क्वेअर वितरण साधारण वितरणाद्वारे अंदाजे केले जाते.



जेव्हा आपण विचार करतो, शून्य अनुमान(null speculation) सत्य आहे, चाचणी आकडेवारीच्या नमुना वितरणास ची-स्क्वेर्ड वितरण म्हणतात. एक किंवा अधिक वर्ग किंवा श्रेणींमध्ये सामान्य फ्रिक्वेन्सी आणि निरीक्षण केलेल्या फ्रिक्वेन्सींमध्ये लक्षणीय फरक आहे की नाही हे निर्धारित करण्यात ची-स्क्वेअर चाचणी मदत करते. हे स्वतंत्र व्हेरिएबल्सची संभाव्यता देते.
टीप: ची-स्क्वॉर्ड चाचणी केवळ लिंग, वय, उंची इत्यादी श्रेणींमध्ये येणारे पुरुष आणि स्त्रिया यासारख्या स्पष्ट डेटासाठी लागू आहे.

5.8.1ची-स्क्वॉर्ड चाचणी(chi-squared test):

ची-स्क्वॉर्ड चाचणी (प्रतिकात्मकरित्या x^2 म्हणून दर्शविली जाते) ही मूलत: व्हेरिएबल्सच्या यादृच्छिक संचाच्या निरीक्षणाच्या आधारे डेटा विश्लेषण असते. सहसा, ही दोन सांख्यिकीय डेटा सेटची तुलना असते. ही चाचणी कार्ल पीअरसन यांनी 1900 मध्ये स्पष्ट डेटा विश्लेषण आणि वितरणासाठी सादर केली होती. म्हणून त्याचा उल्लेख पीअरसनची ची-स्क्वेर्ड चाचणी म्हणून केला गेला.
ची-स्क्वेअर चाचणीचा उपयोग शून्य गृहीतके सत्य मानून, केलेली निरीक्षणे किती संभाव्य असतील याचा अंदाज लावला जातो.
गृहीतक म्हणजे दिलेली स्थिती किंवा विधान सत्य असू शकते असा विचार आहे, ज्याची आपण नंतर चाचणी करू शकतो. ची-स्क्वेअर चाचण्या सामान्यतः स्क्वेअर खोटेपणा किंवा सॅम्पल वेरिएन्सवरील त्रुटींच्या बेरीजमधून तयार केल्या जातात.

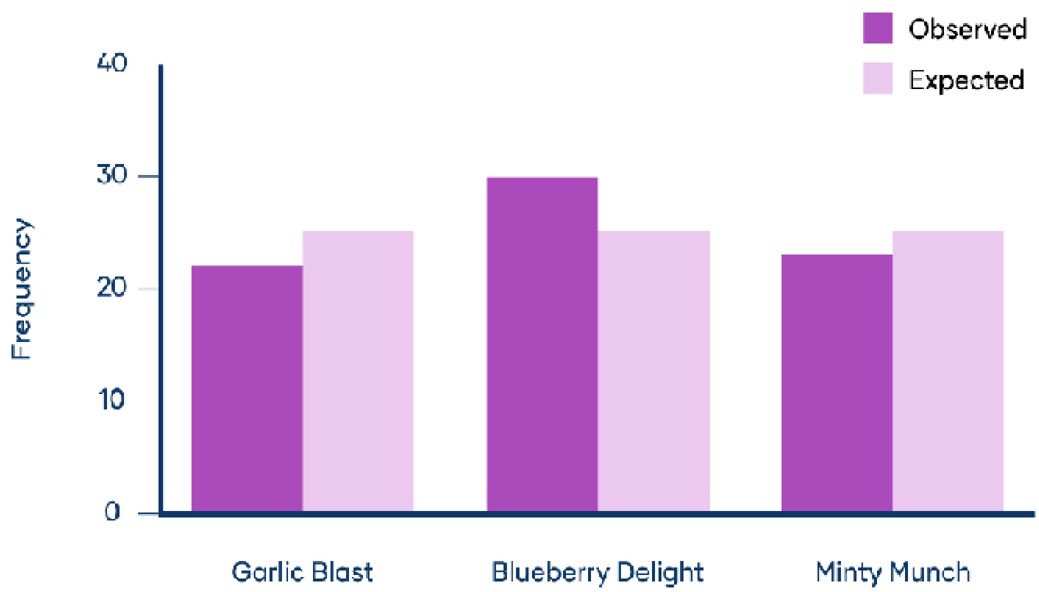
5.9 Degrees of freedom (डिग्री ऑफ फ्रिडम / स्वातंत्र्याचे अंश):
अनुमानात्मक आकडेवारीमध्ये, तुम्ही नमुन्याच्या आकडेवारीची गणना करून लोकसंख्येच्या पॅरामीटरचा अंदाज लावता. आकडेवारीची गणना करण्यासाठी वापरल्या जाणाऱ्या माहितीच्या स्वतंत्र तुकड्यांच्या संख्येला स्वातंत्र्याचे अंश म्हणतात. आकडेवारीच्या स्वातंत्र्याचे प्रमाण नमुन्याच्या आकारावर अवलंबून असते:
जेव्हा नमुन्याचा आकार लहान(small sample) असतो, तेव्हा माहितीचे फक्त काही स्वतंत्र तुकडे असतात आणि म्हणूनच फक्त काही अंश स्वातंत्र्य असते.
जेव्हा नमुन्याचा आकार मोठा(large sample) असतो, तेव्हा माहितीचे अनेक स्वतंत्र तुकडे असतात आणि त्यामुळे स्वातंत्र्याच्या अनेक अंश असतात.
टीप : जरी स्वातंत्र्याचे अंश नमुन्याच्या आकाराशी जवळून संबंधित असले तरी ते समान नाहीत. नमुन्याच्या आकारापेक्षा नेहमीच कमी अंश स्वातंत्र्य असतात.
डिग्री ऑफ फ्रिडम सूत्र:
$D_f = N - 1$
Where
D_f = डिग्री ऑफ फ्रिडम
N = नमुन्याचा आकार
5.14) 20 विद्यार्थ्यांच्या नमुन्यात 10 च्या मानक विचलनासह सरासरी चाचणी स्कोअर 75 आहे. 0.05 स्वातंत्र्याची डिग्री तपासा.
उत्तर:
$D_f = N - 1 = 20 - 1 = 19$
5.15) समजा आपल्याकडे विद्यार्थ्यांचे दोन गट आहेत, गट A (n = 15) आणि गट B (n = 20). दोन गटांसाठी सरासरी गुण खालीलप्रमाणे आहेत:
गट A: मीन = 60, मानक विचलन = 10
गट ब: मीन = 70, मानक विचलन = 12 स्वातंत्र्याची डिग्री तपासा.
उत्तर:
$D_f = N_A + N_B - 2 = 35 - 2 = 33$

5.10 Chi-square goodness of fit test: ची-स्क्वेअर (x^2) तंदुरुस्ती चाचणीची चांगलीता ही एका वर्गीय व्हेरिएबलसाठी फिट चाचणीची चांगलीता आहे. तंदुरुस्तीची चांगलीता हे सांख्यिकीय मॉडेल निरीक्षणांच्या संचाला किती योग्य प्रकारे बसते याचे मोजमाप आहे.
जेव्हा तंदुरुस्तीची चांगलीता जास्त असते, तेव्हा मॉडेलवर आधारित अपेक्षित मूल्ये निरीक्षण केलेल्या मूल्यांच्या जवळ असतात.

जेव्हा तंदुरुस्तपणा कमी असतो, तेव्हा मॉडेलवर आधारित अपेक्षित मूल्ये निरीक्षण केलेल्या मूल्यांपेक्षा खूप दूर असतात.
ची-स्क्वेअर गुडनेस ऑफ फिट चाचण्यांद्वारे विश्लेषित केलेली सांख्यिकीय मॉडेल्स म्हणजे वितरण. ते कोणतेही वितरण असू शकते, सर्व गटांसाठी समान संभाव्यतेच्या सोप्यापासून ते अनेक पॅरामीटर्ससह संभाव्यता वितरणासारखे जटिल.
Example: Observed and expected frequencies
After weeks of hard work, your dog food experiment is complete and you compile your data in a table:

Observed and expected frequencies of cats' flavor choices		
Flavor	Observed	Expected
Garlic Blast	22	25
Blueberry Delight	30	25
Minty Munch	23	25

To help visualize the differences between your observed and expected frequencies, you also create a bar graph:



कॅट फूड कंपनीचे अध्यक्ष तुमचा आलेख पाहतात आणि घोषित करतात की त्यांनी ब्लूबेरी डिलाइटवर लक्ष केंद्रित करण्यासाठी गार्लिक ब्लास्ट आणि मिंटी मंच फ्लेवर्स काढून टाकले पाहिजेत. "खूप वेगाने नको!" तू त्याला सांग.
तुम्ही स्पष्ट करता की तुमची निरीक्षणे तुमच्या अपेक्षेपेक्षा थोडी वेगळी होती, परंतु फरक नाटकीय नाहीत. ते वास्तविक चव प्राधान्याचे परिणाम असू शकतात किंवा ते संधीमुळे असू शकतात.

दुसऱ्या मार्गाने सांगायचे तर: तुमच्याकडे 75 मांजरींचा नमुना आहे, परंतु तुम्हाला खरोखर काय समजायचे आहे ते सर्व मांजरींची लोकसंख्या आहे. हा नमुना मांजरींच्या लोकसंख्येमधून काढला गेला होता जे तिन्ही चव समान वेळा निवडतात?
फिट चाचणी गृहितकांची ची-स्क्वेअर चांगलीता

सर्व गृहितक चाचण्यांप्रमाणे, फिट चाचणीची एक ची-स्क्वेअर चांगलीता दोन गृहितकांचे मूल्यमापन करते: शून्य आणि पर्यायी गृहितके. "निर्दिष्ट वितरणाचे अनुसरण करणाऱ्या लोकसंख्येमधून नमुना काढला होता का?" या प्रश्नाची ते दोन स्पर्धात्मक उत्तरे आहेत.

नल हायपोथिसिस ($\mathbf{H_0}$): लोकसंख्या निर्दिष्ट वितरणाचे अनुसरण करते.

अल्टरनेटीव हायपोथिसिस ($\mathbf{H_a}$): लोकसंख्या निर्दिष्ट वितरणाचे पालन करत नाही.

ही सामान्य गृहीते आहेत जी फिट चाचण्यांच्या सर्व ची-स्क्वेअर चांगुलपणावर लागू होतात. "निर्दिष्ट वितरण" चे वर्णन करून तुम्ही तुमची गृहीते अधिक विशिष्ट बनवावीत. तुम्ही संभाव्यता वितरणाला नाव देऊ शकता (उदा. पॉसॉन वितरण) किंवा प्रत्येक गटाचे अपेक्षित प्रमाण देऊ शकता.

नल हायपोथिसिस ($\mathbf{H_0}$): मांजरीची लोकसंख्या समान प्रमाणात तीन चव निवडते ($\mathbf{P_1} = \mathbf{P_2} = \mathbf{P_3}$).

अल्टरनेटीव हायपोथिसिस ($\mathbf{H_a}$): कुत्र्यांची लोकसंख्या समान प्रमाणात तीन चव निवडत नाही.

Chi-square goodness of fit test formula:

$$x^2 = \sum \frac{(O - E)^2}{E}$$

O is the observed frequency E is the expected frequency

जर x^2 मूल्य गंभीर मूल्यापेक्षा मोठे असेल, तर निरीक्षण केलेल्या आणि अपेक्षित वितरणांमधील फरक सांख्यिकीयदृष्ट्या महत्त्वपूर्ण आहे ($p < \alpha$).

डेटा तुम्हाला शून्य परिकल्पना (नल हायपोथिसिस) नाकारण्याची परवानगी देतो आणि पर्यायी गृहीतकासाठी (अल्टरनेटीव हायपोथिसि)समर्थन प्रदान करतो.

जर x^2 मूल्य गंभीर मूल्यापेक्षा कमी असेल, तर निरीक्षण केलेल्या आणि अपेक्षित वितरणांमधील फरक सांख्यिकीयदृष्ट्या महत्त्वपूर्ण नाही ($p > \alpha$).

डेटा तुम्हाला शून्य गृहीतक ((नल हायपोथिसिस))नाकारण्याची परवानगी देत नाही आणि पर्यायी गृहीतकाला ((अल्टरनेटीव हायपोथिसि))समर्थन देत नाही.

वर्ग परीक्षा निकाल

5.16) समजा करा की एका वर्गामध्ये 100 विद्यार्थ्यांची एक परीक्षा घेतली आहे. या विद्यार्थ्यांना 4 विविध श्रेणींमध्ये (उत्तम, चांगले, सरासरी, आणि निकृष्ट) वर्गीकृत केलं आहे. त्या विद्यार्थ्यांची संख्या खालील प्रमाणे आहे:

उत्तम: 25 विद्यार्थी

चांगले: 30 विद्यार्थी

सरासरी: 35 विद्यार्थी

निकृष्ट: 10 विद्यार्थी

आता आपल्याला हे तपासायचं आहे की हे डेटा एक विशिष्ट वितरण अनुसरण करतं का, म्हणजेच 'समान वितरण' आहे का. Chi-Square चा वापर करून आपण परीक्षण करू शकता.

उत्तर:

सूचना: जर वितरण समान असेल तर प्रत्येक श्रेणीमध्ये सुमारे 25 विद्यार्थी असावेत.

चाचणी सांख्यिकी: Chi-Square तपासणी सूत्र वापरून आपल्याला Chi-Square चा मूल्य काढायचं आहे:

$$x^2 = \sum \frac{(O - E)^2}{E}$$

जिथे:

O_i = निरीक्षित मूल्य (Observed value)

E_i = अपेक्षित मूल्य (Expected value)

निष्कर्ष: Chi-Square मूल्याच्या मदतीने, आपण वितरित वितरणाच्या समानतेवर निर्णय घेऊ शकता. जर Chi-Square मूल्य खूप मोठं आले, तर ह्याचा अर्थ आहे की वर्गीकरण समान नाही.

एक प्रयोगाच्या परिणामांची तुलना

5.17) धरणा करा की एका शाळेत तीन वेगवेगळ्या प्रकारच्या शिकवणी पद्धतींचा अभ्यास करण्यात आलेला आहे. शिकवणी पद्धतींची संख्या आणि परिणाम (स्मार्ट विद्यार्थ्यांची संख्या) खालील प्रमाणे आहे:

पद्धत 1: 10 विद्यार्थ्यांमध्ये 8 स्मार्ट ($O1 = 8$)

पद्धत 2: 10 विद्यार्थ्यांमध्ये 6 स्मार्ट ($O2 = 6$)

पद्धत 3: 10 विद्यार्थ्यांमध्ये 5 स्मार्ट ($O3 = 5$)

उत्तर:

अपेक्षित मूल्ये (Expected values) निकालण्यासाठी, आपण गृहीत धरतो की प्रत्येक पद्धतीतून समान विद्यार्थ्यांना स्मार्ट बनवणे शक्य आहे, म्हणजे प्रत्येक पद्धतीमध्ये 6 विद्यार्थ्यांनाच स्मार्ट होईल ($E = 6$).

Chi-Square चा वापर करून आपण हे तपासू शकता की ती तीन पद्धतीतून मिळालेले डेटा समान आहेत का.

$$x^2 = \sum \frac{(O_1 - E)^2}{E} + \sum \frac{(O_2 - E)^2}{E} + \sum \frac{(O_3 - E)^2}{E}$$

Chi-Square मूल्य मिळवल्यानंतर, आपण विश्लेषण करू शकता की शिकवणी पद्धतींचा प्रभाव समान आहे की नाही.

5.10.1 Chi-square test of independence: ची-स्क्वेअर इंडिपेन्डन्स चाचणी

ची-स्क्वेअर टेस्टचा आणखी एक प्रकार आहे, ज्याला ची-स्क्वेअर टेस्ट ऑफ इंडिपेंडन्स म्हणतात.

जेव्हा तुमच्याकडे एक स्पष्ट व्हेरिएबल असेल आणि तुम्ही त्याच्या वितरणाविषयी गृहीतके तपासू इच्छित असाल तेव्हा फिट चाचणीचा ची-स्क्वेअर चांगुलपणा वापरा.

जेव्हा तुमच्याकडे दोन स्पष्ट व्हेरिएबल्स असतील आणि तुम्हाला त्यांच्या संबंधांबद्दल एक गृहितक तपासायचे असेल तेव्हा स्वातंत्र्याची ची-स्क्वेअर चाचणी वापरा.

चि-स्क्वेअर चाचणी ऑफ इंडिपेंडन्स

5.18)समजा, एका दुकानदाराने दोन प्रकारच्या पेटीमध्ये (पेटी A आणि पेटी B) विक्रीच्या प्रमाणाच्या आधारावर ग्राहकांना दिलेले विक्री देणं तपासायला आहे.

	पेटी A	पेटी B	एकूण
खरीदी	60	40	100
नाही खरीदी	20	30	50
एकूण	80	70	150

चरण 1: नल हायपॉथीसिस:

H0: पेटी प्रकार आणि ग्राहकाची खरेदी स्वतंत्र आहेत.

चरण 2: Expected Frequency (E):

$E = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}$

खरीदी-पेटी A:

$E = \frac{100 \times 80}{150} = 53.33$

खरीदी-पेटी B:

$$E = \frac{100 \times 70}{150} = 46.67$$

नाही खरीदी-पेटी A:

$$E = \frac{50 \times 80}{150} = 26.67$$

नाही खरीदी-पेटी B:

$$E = \frac{50 \times 70}{150} = 23.33$$

चरण 3: Chi-Square (χ^2) गणना:

खरीदी-पेटी A:$\frac{(60-53.33)^2}{53.33} = 0.833$

खरीदी-पेटी B: $\frac{(40-46.67)^2}{46.67} = 0.952$

नाही खरीदी-पेटी A: $\frac{(20-26.67)^2}{26.67} = 1.667$

नाही खरीदी-पेटी B:$\frac{(30-23.33)^2}{23.33} = 1.667$

चि-स्क्वेअर चाचणीची एकूण मूल्य:

χ2=0.833+0.952+1.667+1.667=4.119

चरण 4: Degrees of Freedom (df):

df=(r−1)(c−1)=(2−1)(2−1)=1

प-मूल्य आणि निर्णय:

χ2=4.119 आणि df = 1 असल्यामुळे, p-value साधारणतः 0.042 आहे. 0.05 पेक्षा कमी असल्यामुळे, आपण नल हायपॉथीसिस नाकारतो.

निष्कर्ष:

पेटी प्रकार आणि ग्राहकाची खरेदी एकमेकांशी संबंधित आहेत.

सारांश:

चि-स्क्वेअर चाचणी ऑफ इंडिपेंडन्स दोन वेरिएबल्समधील संबंध तपासते. त्यात observed आणि expected व्हॅल्यूंचा वापर केला जातो. Chi-Square (χ^2) गणना आणि p-मूल्य वापरून आपल्याला नल हायपॉथीसिस स्वीकारायची किंवा नाकारायची असते.

लिंग आणि ऑनलाइन खरेदी करण्याची आवड

5.19)धरणा करा की एका शाळेत १०० विद्यार्थ्यांना दोन गोष्टी विचारल्या:

लिंग: पुरुष (M) किंवा स्त्री (F)

ऑनलाइन खरेदी करण्याची आवड: हो (Y) किंवा नाही (N)

आता आपल्याला तपासायचं आहे की लिंग आणि ऑनलाइन खरेदी करण्याची आवड एकमेकांवर प्रभाव टाकते का, म्हणजे हे दोन व्हेरिएबल्स स्वतंत्र आहेत का?

डेटा:

लिंग / ऑनलाइन खरेदी	हो (Y)	नाही (N)
पुरुष (M)	30	20
स्त्री (F)	25	25

Chi-Square Test for Independence:

दृष्टिकोन: आपण गृहीत धरतो की लिंग आणि ऑनलाइन खरेदीची आवड स्वतंत्र आहेत.

अपेक्षित वारंवारता (Expected Frequency): प्रत्येक सेलसाठी अपेक्षित वारंवारता काढा. यासाठी पुढील सूत्र वापरू शकता:

E = $(\frac{RowTotal)\times(ColumnTotal)}{\text{GrandTotal}})$

Chi-Square चाचणी सूत्र:

$$x^2 = \sum \frac{(O - E)^2}{E}$$

जिथे:

O = निरीक्षित वारंवारता

E = अपेक्षित वारंवारता

गणना:

अपेक्षित वारंवारता काढण्यासाठी:

पुरुषांमध्ये "हो" च्या अपेक्षित वारंवारता: E=(50×55)/100=27.5

पुरुषांमध्ये "नाही" च्या अपेक्षित वारंवारता: E=(50×45)/100=22.5

स्त्रीमध्ये "हो" च्या अपेक्षित वारंवारता: E=(50×55)/100=27.5

स्त्रीमध्ये "नाही" च्या अपेक्षित वारंवारता: E=(50×45)/100=22.5

Chi-Square मूल्य काढण्यासाठी:

$$x^2 = \sum \frac{(30 - 27.5)^2}{27.5} + \sum \frac{(20 - 22.5)^2}{22.5} + \sum \frac{(25 - 27.5)^2}{27.5} + \sum \frac{(25 - 22.5)^2}{22.5}$$

गणना केल्यानंतर आपल्याला Chi-Square मूल्य मिळेल. हे मूल्य तपासणीचे मानक असलेल्या किमतीशी तुलना केली जाऊ शकते.

निष्कर्ष: Chi-Square चाचणीच्या परिणामावरून, आपण निर्णय घेऊ शकता की लिंग आणि ऑनलाइन खरेदीची आवड स्वतंत्र आहेत का.

शिक्षण आणि करियर निवडीचा संबंध

5.20) धरणा करा की एका कंपनीने 100 कर्मचाऱ्यांचा एक सर्व्हे केला आहे. या कर्मचाऱ्यांचे शिक्षण स्तर आणि करियर निवडी यावर आधारित डेटा संकलित केला आहे. आपल्याला तपासायचं आहे की शिक्षण स्तर आणि करियर निवडी यामध्ये कोणताही संबंध आहे का.

डेटा:

शिक्षण स्तर / करियर निवड	व्यवस्थापन (M)	तांत्रिक (T)	विक्री (S)
उच्च शाळा (HS)	5	10	5
पदवी (Bachelors)	15	10	5
पदव्युत्तर (Masters)	10	20	5

Chi-Square Test for Independence:

दृष्टिकोन: गृहीत धरू की शिक्षण स्तर आणि करियर निवडी स्वतंत्र आहेत.
अपेक्षित वारंवारता काढण्यासाठी, प्रत्येक सेलसाठी अपेक्षित वारंवारता काढा.
Chi-Square चाचणी सूत्र वापरून Chi-Square मूल्य काढा.
निष्कर्ष:
Chi-Square चाचणीचा उपयोग करून आपण ठरवू शकता की शिक्षण स्तर आणि करियर निवडी यामध्ये एक दुसऱ्यावर प्रभाव टाकतो का, म्हणजेच ह्या दोन्ही घटकांमध्ये काही प्रमाणात संबंध आहे का.

Contingency tables (आकस्मिक टेबल): जेव्हा तुम्ही स्वातंत्र्याची ची-स्क्वेअर चाचणी करू इच्छित असाल, तेव्हा तुमचा डेटा व्यवस्थित करण्याचा सर्वोत्तम मार्ग म्हणजे फ्रिक्वेंसी वितरण सारणीचा एक प्रकार आहे ज्याला आकस्मिक सारणी म्हणतात.
आकस्मिक सारणी, ज्याला क्रॉस टॅब्युलेशन किंवा क्रॉसटॅब असेही म्हणतात, प्रत्येक गटाच्या संयोजनातील निरीक्षणांची संख्या दर्शवते. यात सहसा पंक्ती(rows) आणि स्तंभ(columns) बेरीज देखील समाविष्ट असतात.
Chi-square test of independence hypothesis (स्वातंत्र्य गृहीतकांची ची-स्क्वेअर चाचणी):
स्वातंत्र्याची ची-स्क्वेअर चाचणी ही एक अनुमानित सांख्यिकीय चाचणी आहे, याचा अर्थ ती तुम्हाला नमुन्याच्या आधारे लोकसंख्येबद्दल निष्कर्ष काढू देते. विशेषत:, हे आपल्याला दोन व्हेरिएबल्स लोकसंख्येमध्ये संबंधित आहेत की नाही हे निष्कर्ष काढू देते.
सर्व गृहीतक चाचण्यांप्रमाणे, स्वातंत्र्याची ची-स्क्वेअर चाचणी शून्य आणि पर्यायी गृहीतकेचे मूल्यांकन करते. गृहितके ही "व्हेरिएबल 1 आणि व्हेरिएबल 2 संबंधित आहेत का?" या प्रश्नाची दोन स्पर्धात्मक उत्तरे आहेत.
नल हायपोथिसिस ($\mathbf{H_0}$): व्हेरिएबल 1 आणि व्हेरिएबल 2 लोकसंख्येमध्ये संबंधित नाहीत; व्हेरिएबल 1 चे प्रमाण व्हेरिएबल 2 च्या भिन्न मूल्यांसाठी समान आहेत.
अल्टरनेटीव हायपोथिसिस ($\mathbf{H_a}$): व्हेरिएबल 1 आणि व्हेरिएबल 2 लोकसंख्येमध्ये संबंधित आहेत; व्हेरिएबल 1 चे प्रमाण व्हेरिएबल 2 च्या भिन्न मूल्यांसाठी समान नाही.
तुम्ही वरील वाक्ये टेम्प्लेट म्हणून वापरू शकता. व्हेरिएबल 1 आणि व्हेरिएबल 2 ला तुमच्या व्हेरिएबल्सच्या नावांसह बदला.

Expected values (अपेक्षित मूल्ये): स्वातंत्र्याची ची-स्क्वेअर चाचणी निरीक्षण आणि अपेक्षित फ्रिक्वेन्सीची तुलना करून कार्य करते. अपेक्षित फ्रिक्वेन्सी अशा असतात की एका व्हेरिएबलचे प्रमाण इतर व्हेरिएबलच्या सर्व मूल्यांसाठी समान असते.
आकस्मिकता सारणी वापरून तुम्ही अपेक्षित फ्रिक्वेन्सी मोजू शकता. पंक्ती r आणि स्तंभ c साठी अपेक्षित वारंवारता आहे:

$$\text{Expected values(E)} = \frac{(Row\ \ total\ \ \ X\, column\ \ total)}{N}$$

Chi-square test of independence formula:

$$x^2 = \sum \frac{(O - E)^2}{E}$$

O is the observed frequency E is the expected frequency

जर x^2 मूल्य गंभीर मूल्यापेक्षा मोठे असेल, तर निरीक्षण केलेल्या आणि अपेक्षित वितरणांमधील फरक सांख्यिकीयदृष्ट्या महत्त्वपूर्ण आहे ($p < \alpha$).
डेटा तुम्हाला व्हेरिएबल्स असंबंधित असल्याच्या शून्य गृहीतकाला नाकारण्याची परवानगी देतो आणि व्हेरिएबल्सशी संबंधित असलेल्या वैकल्पिक गृहीतकाला समर्थन पुरवतो.

जर x^2 मूल्य गंभीर मूल्यापेक्षा कमी असेल, तर निरीक्षण केलेल्या आणि अपेक्षित वितरणांमधील फरक सांख्यिकीयदृष्ट्या महत्त्वपूर्ण नाही ($p > \alpha$).
डेटा तुम्हाला व्हेरिएबल्स असंबंधित असल्याच्या शून्य गृहीतकाला नाकारण्याची परवानगी देत नाही आणि व्हेरिएबल्सशी संबंधित असलेल्या वैकल्पिक गृहीतकाला समर्थन पुरवत नाही.

संदर्भ (References):

1. https://www.coursera.org/learn/r-programming
2. https://www.wolframalpha.com/
3. https://thirdspacelearning.com/gcse-maths/statistics/sampling-methods/

SUGGESTED LEARNING MATERIALS/BOOKS FOR STATISTICAL MODELLING FOR MACHINE LEARNING

Sr No	Author	Title	Publisher
1	Dass H.K Er.Rajnish Verma	Higher Engineering Mathematics	S.Chand Publication,New Delhi ISBN:978-81-2193-890-7
2	S.P Gupta	Statistical methods	S.Chand Publication,New Delhi ISBN:978-93-5161-176-9
3	B.S.Grewal	Higher Engineering Mathematics	Khanna publishers,42nd Edition ISBN :978-81-7409-195-5
4	Dutta .D	A textbook of Engineering Mathematics	New age publications,New Delhi,2006 ISBN:978-21-224-1689-6
5	S.Sampath	Sampling Theory and methods	Narosa publications ISBN: 978-08-493-0980-9