

अध्ययन नोट्स: सांख्यिकी और डेटा विश्लेषण

विषय सूची

1. सांख्यिकी का परिचय
2. केंद्रीय प्रवृत्ति के माप
3. फैलाव के माप
4. डेटा के प्रकार
5. संभाव्यता और वितरण
6. नमूनाकरण और अनुमान
7. परिकल्पना परीक्षण
8. सहसंबंध और प्रतिगमन
9. सारांश और मुख्य सूत्र

1. सांख्यिकी का परिचय

◆ सांख्यिकी क्या है?

- वर्णनात्मक सांख्यिकी: डेटा को सारांशित और वर्णित करती है।
- अनुमानात्मक सांख्यिकी: नमूना डेटा के आधार पर जनसंख्या के बारे में पूर्वानुमान और अनुमान लगाती है।

2. केंद्रीय प्रवृत्ति के माप

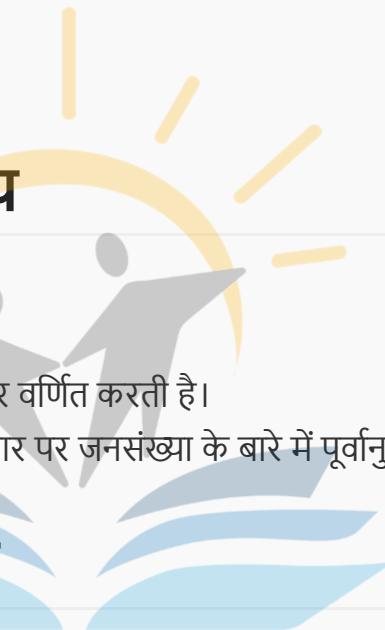
◆ माध्य, माध्यिका और बहुलक

- **माध्य (अंकगणितीय माध्य)**: सभी मानों का औसत।
- $\bar{x} = \frac{\sum x_i}{n}$
- **माध्यिका**: क्रमबद्ध डेटा का मध्य मान।
- **बहुलक**: सबसे अधिक बार आने वाला मान।

◆ भारित माध्य

- जब मानों के अलग-अलग वजन हों।

$$\bar{x}_W = \frac{\sum w_i x_i}{\sum w_i}$$



◆ गुणोत्तर माध्य

- विकास दर या प्रतिफल दर की गणना के लिए उपयोग किया जाता है।
- $GM = (\prod x_i)^{1/n}$

◆ हरात्मक माध्य

- दरों और अनुपातों के लिए प्रयुक्त।

$$HM = \frac{n}{\sum \frac{1}{x_i}}$$

◆ केंद्रीय प्रवृत्ति के मापों की तुलना

माप	परिभाषा	उपयोग के मामले
माध्य	सभी मानों का औसत	सतत डेटा
माध्यिका	क्रमबद्ध डेटा का मध्य मान	विषम डेटा या बाहरी मान
बहुलक	सबसे अधिक आने वाला मान	श्रेणीबद्ध डेटा
भारित माध्य	भार के आधार पर समायोजित औसत	वित्तीय विश्लेषण, भारित डेटा

◆ 3. फैलाव के माप

◆ परिसर

- परिभाषा:** अधिकतम और न्यूनतम मानों के बीच का अंतर।
- $R = \text{Max} - \text{Min}$

◆ प्रसरण और मानक विचलन

- प्रसरण:** माध्य से वर्गित विचलनों का औसत।
- $\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$
- मानक विचलन:** प्रसरण का वर्गमूल।
- $\sigma = \sqrt{\sigma^2}$

◆ परिवर्तन का गुणांक

- परिभाषा:** मानक विचलन को माध्य से विभाजित किया जाता है।
- $CV = \frac{\sigma}{\mu} \times 100\%$

- **उपयोग:** डेटासेट के बीच परिवर्तनशीलता की तुलना करने के लिए।

◆ चतुर्थक, दशमक और प्रतिशतक

माप	विवरण
चतुर्थक	डेटा को चार बराबर भागों में विभाजित करती है
दशमक	डेटा को दस बराबर भागों में विभाजित करती है
प्रतिशतक	डेटा को 100 बराबर भागों में विभाजित करती है

📌 4. डेटा के प्रकार

◆ नाममात्र डेटा

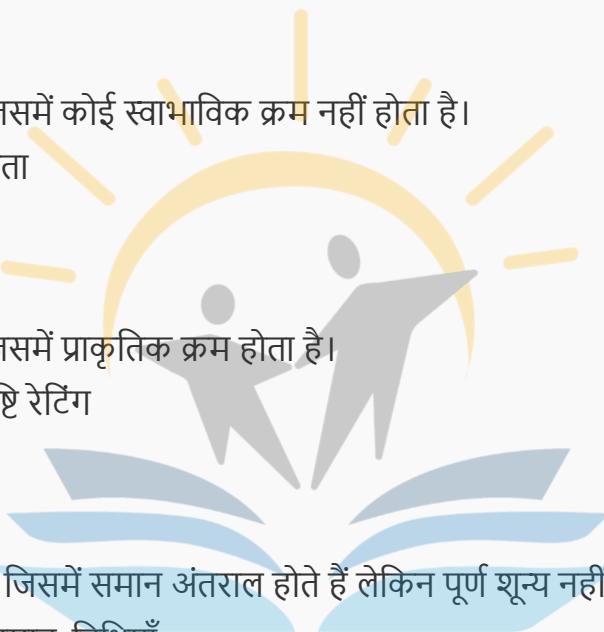
- **परिभाषा:** श्रेणीबद्ध डेटा जिसमें कोई स्वाभाविक क्रम नहीं होता है।
- **उदाहरण:** लिंग, रंग, राष्ट्रियता

◆ क्रमसूचक डेटा

- **परिभाषा:** श्रेणीबद्ध डेटा जिसमें प्राकृतिक क्रम होता है।
- **उदाहरण:** शिक्षा स्तर, संतुष्टि रेटिंग

◆ अंतराल डेटा

- **परिभाषा:** संख्यात्मक डेटा जिसमें समान अंतराल होते हैं लेकिन पूर्ण शून्य नहीं होता है।
- **उदाहरण:** सेल्सियस में तापमान, तिथियाँ



◆ अनुपात डेटा

- **परिभाषा:** संख्यात्मक डेटा जिसमें समान अंतराल और पूर्ण शून्य होता है।
- **उदाहरण:** ऊँचाई, वजन, आय

📌 5. संभाव्यता और वितरण

◆ मूल संभाव्यता

- **संभाव्यता:** किसी घटना के घटित होने की संभावना।
- $P(A) = \frac{\text{अनुकूल परिणामों की संख्या}}{\text{कुल परिणामों की संख्या}}$

◆ असतत बनाम सतत वितरण

प्रकार	विवरण	उदाहरण
असतत	विशिष्ट, अलग-अलग मान लेता है	द्विपद, पॉयसन, बनौली
सतत	किसी सीमा में कोई भी मान लेता है	सामान्य, घातांकीय, एकसमान

◆ सामान्य वितरण

वितरण	विवरण	सूत्र
द्विपद	सफलता/विफलता परिणाम	$P(X = k) = \binom{n}{k} p^k (1 - p)^{n - k}$
पॉयसन	समय के साथ दुर्लभ घटनाएँ	$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$
सामान्य	सममित वितरण	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x - \mu)^2}{2\sigma^2}}$

◆ 6. नमूनाकरण और अनुमान

◆ नमूनाकरण विधियाँ

विधि	विवरण
सरल यादृच्छिक नमूनाकरण	प्रत्येक सदस्य के चयन की समान संभावना
स्तरीकृत नमूनाकरण	जनसंख्या को स्तरों में विभाजित किया जाता है
समूह नमूनाकरण	तत्वों के समूह चुने जाते हैं
व्यवस्थित नमूनाकरण	नियमित अंतराल पर तत्व चुने जाते हैं

◆ बिंदु अनुमान

- परिभाषा:** जनसंख्या पैरामीटर के अनुमान के लिए एकल मान।
- उदाहरण:** नमूना माध्य, नमूना अनुपात

◆ अंतराल अनुमान

- परिभाषा:** मानों की एक सीमा जिसमें जनसंख्या पैरामीटर के शामिल होने की संभावना होती है।
- विश्वास अंतराल:**

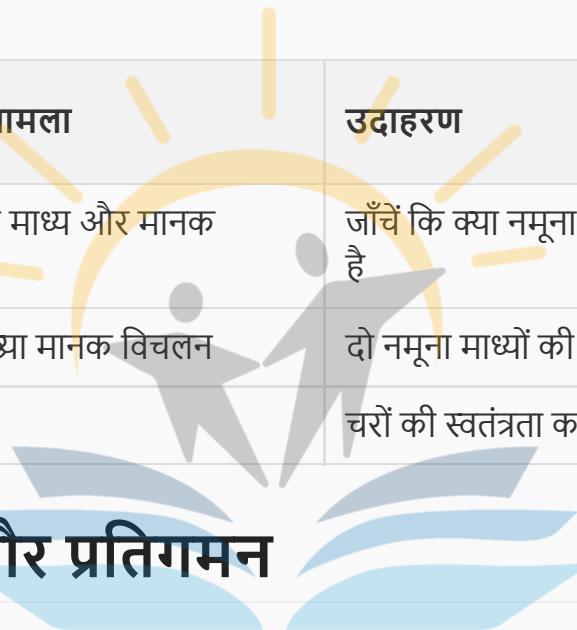
- $\bar{x} \pm z \cdot \frac{\sigma}{\sqrt{n}}$

📌 7. परिकल्पना परीक्षण

◆ परिकल्पना परीक्षण के चरण

1. परिकल्पनाओं को बताएँ (शून्य और वैकल्पिक)
2. सार्थकता स्तर चुनें (α)
3. परीक्षण सांख्यिकी की गणना करें
4. महत्वपूर्ण क्षेत्र निर्धारित करें
5. निर्णय लें

◆ सामान्य परीक्षण



परीक्षण प्रकार	उपयोग का मामला	उदाहरण
Z-परीक्षण	ज्ञात जनसंख्या माध्य और मानक विचलन	जाँचें कि क्या नमूना माध्य जनसंख्या माध्य से भिन्न है
T-परीक्षण	अज्ञात जनसंख्या मानक विचलन	दो नमूना माध्यों की तुलना
ची-स्कायर	श्रेणीबद्ध डेटा	चरों की स्वतंत्रता का परीक्षण

📌 8. सहसंबंध और प्रतिगमन

SATHEE

◆ सहसंबंध

- **परिभाषा:** एक रैखिक संबंध की शक्ति और दिशा का माप।
- **पियर्सन का सहसंबंध गुणांक:**

$$\bullet \quad r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

◆ प्रतिगमन विश्लेषण

- सरल रैखिक प्रतिगमन:
- $y = a + bx$
- $b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$
- $a = \bar{y} - b\bar{x}$

📌 9. सारांश और मुख्य सूत्र

विषय	सूत्र
माध्य	$\bar{x} = \frac{\sum x_i}{n}$
भारित माध्य	$\bar{x}_W = \frac{\sum w_i x_i}{\sum w_i}$
गुणोत्तर माध्य	$GM = (\prod x_i)^{1/n}$
हरात्मक माध्य	$HM = \frac{n}{\sum \frac{1}{x_i}}$
प्रसरण	$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$
मानक विचलन	$\sigma = \sqrt{\sigma^2}$
परिवर्तन का गुणांक	$CV = \frac{\sigma}{\mu} \times 100\%$
द्विपद संभाव्यता	$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$
सामान्य वितरण	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

📌 निष्कर्ष

SATHEE

सांख्यिकी डेटा को समझने और व्याख्या करने के लिए एक शक्तिशाली उपकरण है। केंद्रीय प्रवृत्ति, फैलाव, संभाव्यता और परिकल्पना परीक्षण जैसी प्रमुख अवधारणाओं में महारत हासिल करके आप डेटा से सूचित निर्णय ले सकते हैं और सार्थक निष्कर्ष निकाल सकते हैं।

📌 अंतिम नोट्स

- समझ को मजबूत करने के लिए अभ्यास समस्याएँ।
- डेटा विश्लेषण के लिए एक्सेल, R या पायथन जैसे सॉफ्टवेयर टूल का उपयोग करें।
- मुख्य सूत्रों और अवधारणाओं को बनाए रखने के लिए नियमित रूप से समीक्षा करें।