स्टेटिस्टिक्स फॉर डाटा साइंस**- 1**

**प्रो. उषा मोहन**

**प्रबंधन अध्ययन विभाग**

**भारतीय प्रौद्योगिकी संस्थान, मद्रास**

**लैक्चर -01**

**इंट्रोडक्शन**

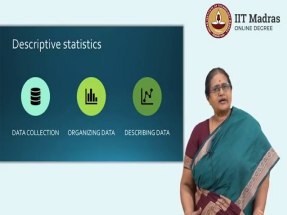
**नमस्कार। पाठ्यक्रम स्टेटिस्टिक्स फॉर डाटा साइंस-1 में आपका स्वागत है। मैं उषा मोहन, प्रबंधन अध्ययन विभाग, आईआईटी मद्रास में प्रोफेसर हूं। अगले 12 हफ्तों में हम साथ मे आपके ऑनलाइन डिग्री प्रोग्राम के लिए आवश्यक स्टेटिस्टिक्स की मूल बातें समझने के लिए अध्यन्न करेंगे। तो आइये हम सबसे पहले यह समझने से शुरूआत करते हैं कि स्टेटिस्टिक्स या सांख्यिकी क्या है।**

**(स्लाइड समय देखें: 00:41)**



**तो आइये सबसे पहले हम यह समझते हैं कि सांख्यिकी की परिभाषा क्या है। सांख्यिकी की सबसे लोकप्रिय परिभाषाओं में से एक यह है की सांख्यिकी डाटा से सीखने की एक कला है जिसमें शामिल है, डाटा का संग्रह, डाटा का विश्लेषण, डाटा की प्रस्तुति और फिर अंत मे डाटा से निष्कर्ष निकालना। इस पाठ्यक्रम में आप जो मुख्य बातें सीखने जा रहे हैं, उनमें सबसे पहले यह है कि हम ये समझेंगे कि डाटा कैसे एकत्र किया जाता हैं। जी हाँ यह मुख्य बातों में से एक है, और हम इस से ही शुरुआत करेंगे।**

**(स्लाइड समय देखें: 01:16)**



**तो, हम डाटा संग्रह के साथ शुरू करते हैं। सबसे पहले हम यह देखेंगे कि आप डाटा कैसे एकत्र करते हैं और फिर हम यह देखेंगे कि आप इस डाटा को कैसे प्रस्तुत करते हैं। इस पाठ्यक्रम के दौरान हम यह सीखेंगे कि आप डाटा को सारणीबद्ध रूप (tabular form) में कैसे प्रस्तुत करते हैं; यहाँ हम frequency tables यानि बारंबारता सारणियों की अवधारणा का आपसे परिचय कराएंगे साथ ही हम आपेक्षिक आवृत्ति की अवधारणा यानि relative frequency का भी आपसे परिचय कराएंगे और फिर हम देखेंगे कि आप डाटा को कैसे सारणीबद्ध करते हैं। हम यहाँ विभिन्न प्रकार के डाटा के बारे में भी सीखेंगे।**

**इसके बाद हम डाटा का वर्णन ग्राफ़िकल तकनीकों का उपयोग करके देखेंगे। जो ग्राफिकल तकनीकें आप यहाँ सीखेंगे, वे हैं बार चार्ट, पाई चार्ट, हिस्टोग्राम, स्टेम लीफ प्लॉट वगैरह वगैरह । फिर हम यहाँ यह भी सीखेंगे कि कैसे आप डाटा के संख्यात्मक माप प्राप्त कर सकते हैं। इसमे हम यहां मुख्य रूप से central tendency यानि केंद्रीय प्रवृत्ति के मापों को अर्थात्, माध्य (mean), माध्यिका (median), बहुलक (mode) और विचरण के माप (measure of variation) अर्थात्, श्रेणी (range), विचरण (variance) और मानक विचलन (standard deviation) को भी समझेंगे।**

.

**अभी तक हमने केवल एक चर (variable) का वर्णन किया है, अब हम यहाँ चरों के बीच संबंध यानि variables के बीच संबंध के बारे में भी विचार करेंगे। इसके लिए हम यहाँ आपको contingency tables यानि आकस्मिक तालिकाओं, स्कैटर प्लॉट्स की अवधारणा से अवगत कराएंगे साथ ही साथ हम आपको यहाँ सहसंबंध मैट्रिक्स (correlation matrix) से भी अवगत कराएंगे। इसलिए, मुख्य रूप से इस वर्णनात्मक सांख्यिकी मॉड्यूल में आप डाटा एकत्र करने, डाटा व्यवस्थित करने और डाटा का वर्णन करने के बारे में सीखेंगे। डाटा का वर्णन हम यहाँ ग्राफिकल तकनीकों और संख्यात्मक तकनीकों दोनों का उपयोग करके समझेंगे।**

**जब आप डाटा का वर्णन करने में खुद को सहज पाते है और जो भी डाटा आपके पास है उसका वर्णन आप करना चाहते हैं, तो जैसे की मैंने आपको पहले ही बताया की सांख्यिकी, डाटा से सीखने की एक कला है, आपका अगला कदम अब यह होगा की आप डाटा से निष्कर्ष निकालने की कोशिश करे। जी हाँ यही आपका अगला कदम होगा जहां आप डाटा से कोई निष्कर्ष निकालने की कोशिश करेंगे। इसी कोशिश के साथ आप इस विषय के अगले स्तर पर पहुचेंगे जो की होगा inferential statistics (अनुमानात्मक आंकड़े)। inferential statistics के बारे मे आप सांख्यिकी के अगले पाठ्यक्रम में पढ़ेंगे। हालांकि आपकी सुविधा के लिए इस पाठ्यक्रम में भी हम आपको inferential statistics को समझने के लिए नींव तैयार कराएंगे।**

**inferential statistics यानि अनुमानित आंकड़ों को समझने के लिए नींव तैयार करने से मेरा क्या तात्पर्य है? इसकी उत्सुकता को बढ़ाने के लिए हम probability theory (संभाव्यता सिद्धांत) के संछिप्त परिचय के माध्यम से करेंगे।**

**(स्लाइड समय देखें: 04:09)**



.

**जब हम probability theory (संभाव्यता सिद्धांत) की बात करते है तो हम permutation और combination यानि क्रमपरिवर्तन और संयोजनों की बात करते हैं। आपमे से बहुत से लोगों ने इनके बारे मे अपने हाई स्कूल में सीखा ही होगा। यहाँ आप permutation और combination के बारे मे एक बार फिर विचार करने जा रहे हैं। इसके पश्चात हम रैनडम एक्सपेरीमेंट्स (यादृच्छिक प्रयोगों) के माध्यम से probability (संभाव्यता) की धारणा का आपसे परिचय कराएंगे जिसमे आप rolling dice, tossing coin (एक पासा लुढ़कना, एक सिक्का उछालना) के बारे मे समझेंगे। फिर हम घटनाओं की संभावना, घटनाओं के संयोजन और वे क्यों आवश्यक हैं के बारे में बात करेंगे।**

**इस पाठ्यक्रम के अंत मे हम आपको को दो मुख्य वितरणों से परिचित कराएंगे अर्थात्, द्विपद वितरण और सामान्य वितरण (binomial distribution and nominal distribution)। पूरे पाठ्यक्रम में, वैचारिक स्तर पर शिक्षण और इन अवधारणाओं को वास्तविक दुनिया की समस्याओं पर लागू करने पर ध्यान केंद्रित किया जाएगा। इसके लिए आपको बहुत सारे असाइनमेंट करने होंगे जो आपके ज्ञान का परीक्षण वैचारिक (conceptual) स्तर और एप्लिकेशन स्तर पर करेंगे।**

**पाठ्यक्रम के हर स्तर पर हमारी यही कोशिश होगी की शिक्षार्थी को, इस पाठ्यक्रम के लिए विशेष रूप से तैयार किए गए केस स्टडीज के सेट के माध्यम से, अवधारणाओं को समझने और उन अवधारणाओं को लागू करने के लिए प्रेरित किया जा सके।**

**सांख्यिकी की दुनिया और इस पाठ्यक्रम,** स्टेटिस्टिक्स फॉर डाटा साइंस**- 1 मे आपका फिर से स्वागत है।**

**GLOSSARY**

|  |  |  |
| --- | --- | --- |
| **ENGLISH WORD** | **HINDI WORD** | **MEANING** |
| **Statistics** | **स्टेटिस्टिक्स** | **सांख्यिकी** |
| **Tabular Form** | **टबुलर फॉर्म** | **सारणीबद्ध रूप** |
| **Frequency Table** | **फ्रीकुएनसी टेबल** | **बारंबारता सारणियों** |
| **Relative Frequency** | **रिलेटिव फ्रीकुएनसी** | **आपेक्षिक आवृत्ति** |