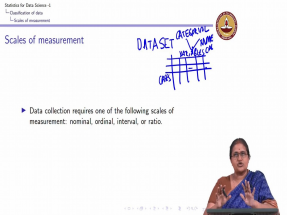
स्टेटिस्टिक्स फॉर डाटा साइंस- 1 प्रो. उषा मोहन प्रबंधन अध्ययन विभाग भारतीय प्रौद्योगिकी संस्थान, मद्रास

लेक्चर - ०६

इंट्रोडक्शन एंड टाइप्स ऑफ़ डाटा पार्ट – ४

(स्लाइड समय देखें: 00:16)



पाठ्यक्रम में इस स्तर पर आ कर, मेरा मानना है की अब आप अच्छे से जानते हैं कि डाटा क्या है, और मेरा यह भी मानना है की आपके पास एक अच्छा डाटा सेट उपलब्ध है। अच्छे डाटा सेट से मेरा मतलब है कि आपके पास वास्तव में एक डाटा सेट है जो वेरिएबल (चर) में व्यवस्थित है और केसेस के रूप मे प्रदर्शित है। प्रत्येक चर के लिए प्रत्येक केस दर्ज किया गया है और यदि हमारे पास कोई डाटा उपलब्ध नहीं है, तो हम उस डाटा को रिकॉर्ड नहीं करेंगे। तो, यह एक सुव्यवस्थित डाटा सेट है।

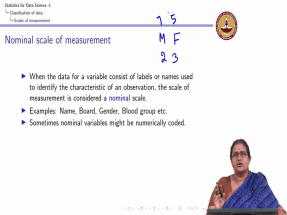
इसके अलावा, इस डाटा सेट से आपको इन चरों को श्रेणीबद्ध चर के रूप में तथा संख्यात्मक चर के रूप में वर्गीकृत करने में भी सक्षम होने चाहिए। तो इस स्तर तक आपको अपने डाटा सेट के साथ ऐसे वर्गिकरण करने में सक्षम हो जाना चाहिए।

तो अब हम आगे क्या करना चाहते हैं? अब हम फिर से अपने प्रत्येक वेरिएबल (चर) की अधिक विस्तार से जांच करेंगे क्योंकि यदि आप मेरे द्वारा पहले दी गयी स्टेटिस्टिक्स की परिभाषा को देखेंगे जहां यह कहा गया था कि स्टेटिस्टिक्स, डाटा से सीखने की काला हैं तो जब हम डाटा से सीखने की बात करते हैं तो तत्काल यह देखना चाहते हैं कि क्या हम अपने डाटा को सारांशित कर सकते हैं। यहाँ हम दोहराना चाहेंगे की जब हम डाटा को सारांशित करने के बारे मे कहते हैं तो जो प्रश्न हमारे सामने आता है वह यह है कि क्या हम डाटा से कुछ ग्राफिकल सारांश या संख्यात्मक (न्यूमेरिकल) सारांश प्राप्त कर सकते हैं। जिस क्षण, हम संख्यात्मक यानि न्यूमेरिकल सारांश की बात करते हैं हम यह जानने की कोशिश करते हैं कि क्या हम डाटा पर अंकगणितीय संचालन (arithmetic operations) कर सकते हैं या नहीं, ठीक है।

तो यह जानने के लिए कि क्या हम डाटा पर अंकगणितीय संचालन कर सकते हैं, हमे यह भी समझने की जरूरत होगी कि अपने डाटा के लिए हम कौन सी मेज़रमेंट स्केल (माप के पैमाने) का उपयोग कर सकते हैं। जब हम माप के पैमाने को देखते हैं, तो हमारे पास माप के 4 पैमाने हो सकते हैं, उन्हें नाममात्र (nominal), क्रमिक (ordinal), अंतराल (interval) और अनुपात (ratio) स्केल या पैमाना कहा जाता है। हम माप के इन पैमानों के बारे में आगे विस्तार से समझेंगे।

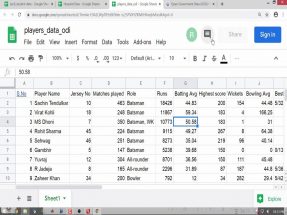
यह समझना आखिरकार महत्वपूर्ण क्यों है? हमारे लिए यह जानना अत्यंत महत्वपूर्ण है कि हमारे डाटा सेट में मौजूद प्रत्येक चर के लिए माप का पैमाना क्या है जिससे अंततः यह जान सकें की हम उस चर के लिए किस प्रकार का सारांश हांसिल कर सकते हैं। इसलिए, हमारे लिए यह जानना अत्यंत महत्वपूर्ण है कि प्रत्येक चर के लिए माप का पैमाना क्या है।

(स्लाइड समय देखें: 03:01)

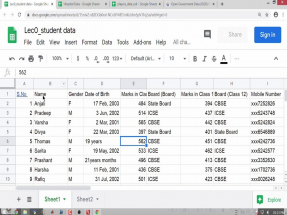


तो हम नॉमिनल स्केल (nominal scale) से शुरू करते हैं। जब डाटा में लेबल या नाम होते हैं, तो माप के पैमाने को नॉमिनल स्केल (nominal scale) माना जाता है।

(स्लाइड समय देखें: 03:16)

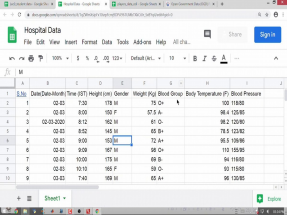


(स्लाइड समय देखें: 03:19)



तो आइए हम अपने पिछले उदाहरण पर वापस चलते हैं। आप पहले उदाहरण में देख सकते हैं कि हमारे पास नाम है जो स्पष्ट रूप से नॉमिनल (नाममात्र का पैमाना) है क्योंकि यह केवल नाम है, ठीक है। फिर आप देख सकते हैं की यहाँ बोर्ड है जो लेबल ही है, हमारे पास राज्य बोर्ड है हमारे पास ICSE (आईसीएसई) बोर्ड है, हमारे पास यहाँ सीबीएसई बोर्ड भी है, ठीक है। जेंडर भी यहाँ एक लेबल ही है जिसे हम यहाँ एक श्रेणी के तौर पर महिला और पुरुष के रूप मे ले सकते हैं।

(स्लाइड समय देखें: 03:57)



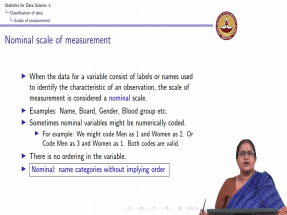
तो, आप देख सकते हैं हमारे पास डाटा के तौर पर नाम, बोर्ड और जेंडर हैं। हमारे पास ब्लड ग्रुप का डाटा है। रक्त समूह डाटा में आप फिर से देख सकते हैं कि हम इसे ओ पॉजिटिव, ए माइनस, ओ नेगेटिव के रूप में लेबल कर सकते हैं। एक बात हमें यहां ध्यान देने की जरूरत है कि हमारे पास यहाँ सिर्फ लेबल हैं, हमारे पास यहाँ केवल नाम हैं और इन नामों का कोई विशेष क्रम भी नहीं है।

उदाहरण के लिए डाटा से कोई फर्क नहीं पड़ता कि हम पहले महिला फिर पुरुष ले या पहले पुरुष फिर महिला लिखे। हम इसे विशेषता की पहचान करने के लिए सिर्फ एक लेबल के रूप में उपयोग कर रहे हैं। इसे माप का नाममात्र पैमाना यानि नॉमिनल स्केल ऑफ मेज़रमेंट कहा जाता है। कभी-कभी इन नॉमिनल वेरिएबल को संख्यात्मक रूप से कोडेड किया जा सकता है। संख्यात्मक रूप से कोडेड नॉमिनल वेरिएबल से हमारा क्या मतलब है?

उदाहरण के लिए हमारे पास जेंडर है। जेंडर दो लेबल ले सकता है - नर (male) और मादा (female) है। हम पुरुष को शून्य (0) और महिला को एक (1) कोड दे सकते हैं, हम पुरुष को एक (1) और महिला को शून्य (0) कोड भी दे सकते हैं। यह संख्यात्मक रूप से कोड देना वेरिएबल को लेबल करने के ही बराबर है, यहाँ संख्यात्मक; 0 या 1 होने के बारे में कोई निश्चितता नहीं है।

हम एक पुरुष को 2 और महिला को 3, या महिला को 5 और पुरुष को 7 भी लेबल कर सकते हैं। यह 0 या 1 या कोई भी नंबर की कोडिंग केवल लेबल को समझने के लिए ही है; जब हम नॉमिनल वेरिएबल को कोड कर रहे होते हैं तो इन संख्याओं का कोई अर्थ नहीं होता है। दोनों ही कोड मान्य हैं, यहाँ हमारा यही मतलब है।

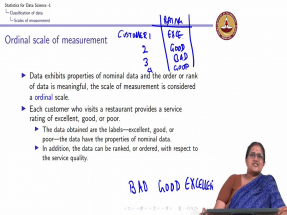
(स्लाइड समय देखें: 05:40)



जब नॉमिनल वेरिएबल को 1 या 2 या फिर 3 कोड नहीं दिया जाता है तो यह दोनों ही कोड मान्य हैं। यहाँ कोई क्रम नहीं है। यह अत्यंत महत्वपूर्ण है कि हम यह समझे की जब हम नॉमिनल वेरिएबल के बारे में बात करते हैं तो वेरिएबल में कोई क्रम नहीं होता है। यानि नॉमिनल वेरिएबल केवल नाम श्रेणियां हैं जो किसी क्रम को नहीं दर्शाती हैं।

जब हम अपने छात्रों के डाटा सेट पर वापस जाते हैं तो वहाँ नाम, जेंडर, बोर्ड नॉमिनल वेरिएबल हैं। ब्लड बैंक डाटा सेट में हमारे पास जेंडर और ब्लड ग्रुप है जो नॉमिनल वेरिएबल हैं। क्रिकेटिंग डाटा सेट में जर्सी नंबर नॉमिनल वेरिएबल है। बल्लेबाज की भूमिका भी नॉमिनल वेरिएबल है, साथ ही खिलाड़ी का नाम भी नॉमिनल वेरिएबल है। तो नॉमिनल स्केल ऑफ मेज़रमेंट यानि माप के नाममात्र पैमाने का उपयोग तब किया जाता है जब हमारे पास केवल नाम श्रेणियां हों जो किसी क्रम को नहीं दर्शाती हों।

(स्लाइड समय देखें: 06:55)

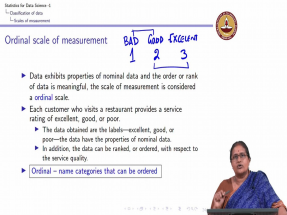


माप के अगले पैमाने को माप का क्रमिक पैमाना कहा जाता है यानि ओर्डिनल स्केल ऑफ मेज़रमेंट जहां डाटा नॉमिनल डाटा के समान ही गुण को प्रदर्शित करता है, लेकिन यहां एक क्रम या रैंक की उपस्थित होती हैं। इससे हमारा क्या मतलब है समझने के लिए उदाहरण के तौर पर एक ग्राहक जो किसी रेस्तरां में जाता है, वह वहाँ उत्कृष्ट, अच्छी और बुरी तीन स्तर पर सेवा की रेटिंग प्रदान करता है तो यह डाटा लेबल ही हैं।

ऊपर दिये उदाहरण मे डाटा फिर से लेबल ही है। उदाहरण के लिए जैसे हमारे पास डाटा हो सकता है जहां ग्राहक 1 है जो उत्कृष्ट रेटिंग देता है, ग्राहक 2 अच्छी सेवा की रेटिंग देता हैं तथा ग्राहक 3 खराब या बुरी सेवा की रेटिंग देता है, ग्राहक 4 फिर से सेवा को अच्छी सेवा के तौर पर रेटिंग देता हैं। तो यदि आप वेरिएबल को देखें, तो यहाँ वेरिएबल रेटिंग है।

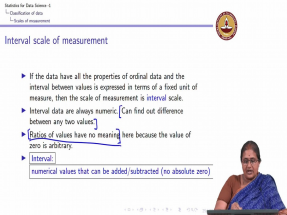
यहां आप फिर से देख सकते हैं कि यह नॉमिनल वैल्यू ले रहा है। नॉमिनल से हमारा तात्पर्य यह हैं की यह एक श्रेणीगत वैल्यू ले रहा है जहां हमारे पास श्रेणीगत वेरिएबल में 3 श्रेणियां हैं; उत्कृष्ट, अच्छी और बुरी, लेकिन इस श्रेणीगत चर के भीतर एक क्रम है। आप जानते हैं कि क्रम बुरी सेवा, अच्छी सेवा और उत्कृष्ट सेवा है। तो श्रेणीबद्ध या नॉमिनल डाटा जो किसी रैंक या क्रम को प्रदर्शित करता है उसे माप का क्रमिक पैमाना या ओर्डिनल स्केल ऑफ मेज़रमेंट कहा जाता है।

(स्लाइड समय देखें: 08:42)



तो ओर्डिनल स्केल ऑफ मेज़रमेंट डाटा वो नाम श्रेणियां हैं जिन्हें क्रमित किया जा सकता है।

(स्लाइड समय देखें: 08:53)



अगले मेज़रमेंट स्केल को इंटरवल मेज़रमेंट स्केल कहा जाता है। जब हम क्रमिक पैमाने यानि ओर्डिनल स्केल के बारे में बात करते हैं तो हम माप के क्रमिक पैमाने को कोड कर सकते हैं। उदाहरण के लिए हमारे पहले के उदाहरण में खराब सेवा को 1 के रूप में कोड किया जा सकता था, अच्छी सेवा को 2 के रूप में कोड किया जा सकता हैं और उत्कृष्ट सेवा को 3 के रूप में कोड किया जा सकता था। हम उन्हें इस तरह से कोड कर सकते थे।

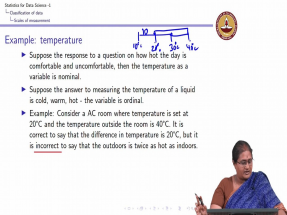
यहाँ १, २, ३ कोड में क्रम है। लेकिन फिर एक बात जो हमे यहां समझने की जरूरत है वह यह है कि बुरी सेवा से अच्छी सेवा के बीच का अंतर या अच्छी सेवा और उत्कृष्ट सेवा के बीच का अंतर समान नहीं होगा। यह सिर्फ एक क्रम है, हमे पता है कि उत्कृष्ट सेवा , अच्छी सेवा से बेहतर है लेकिन हम यह नहीं कह सकते कि उत्कृष्ट सेवा और अच्छी सेवा के बीच का अंतर उतना ही है जितना कि अच्छी और बुरी सेवा के बीच का अंतर। हमारे पास यहाँ एक क्रम है लेकिन इस समय हम इस क्रम के बारे में और कुछ भी टिप्पणी करने में सक्षम नहीं हैं।

तो जब हम इंटरवल मेज़रमेंट स्केल की बात करते हैं तो इंटरवल स्केल के डाटा मे अंतराल पैमाने (interval scale of data) के सभी गुण होते हैं, लेकिन मानों के बीच का अंतराल माप की एक निश्चित इकाई को व्यक्त करता है। आपको याद है जब हमने कहा की सेवा बुरी, अच्छी और उत्कृष्ट हैं, तो हमने यह भी कहा कि अच्छे और बुरे के बीच का अंतर ज़रूरी नहीं की उतना ही नहीं होना चाहिए जितना कि उत्कृष्ट और अच्छे के बीच का अंतर।

जबकि जब हमारे पास इंटरवल डाटा होता है तो हमारे सामने एक ऑर्डरिंग होती है, लेकिन उस मामले में जब भी हम अपने डाटा को ऑर्डर कर रहे होते हैं तो मानों के बीच का अंतराल माप की एक निश्चित इकाई में व्यक्त किया जाता है। जब माप की एक निश्चित इकाई में अंतर व्यक्त किया जाता है तो हम किन्हीं दो मानों के बीच के अंतर का पता लगा सकते हैं, ठीक है।

अब अगर हम रैशियो स्केल की बात करे तो यहाँ अनुपातों (ratios) का कोई अर्थ नहीं है क्योंकि शून्य (0) का मान मनमाना है। आइए इसे एक उदाहरण के माध्यम से समझते हैं। अंतराल डाटा (interval data) और संख्यात्मक मान (numerical value) जिन्हें जोड़ा या घटाया जा सकता है, उनका कोई निरपेक्ष शून्य (0) नहीं है।

(स्लाइड समय देखें: 11:05)



आइए एक उदाहरण के रूप में तापमान को ले। मान लीजिए, किसी प्रश्न का संदर्भ या प्रतिक्रिया यह है कि दिन कितना गर्म है, और आप प्रतिक्रिया को केवल आरामदायक, असहज या सिर्फ अच्छा या बुरा कहते हैं तो हम सिर्फ अपनी भावना को एक लेबल दे रहे हैं। हम वास्तव में अपनी भावनाओं की ग्रेडिंग या ऑर्डरिंग नहीं कर रहे है, चाहे हम सिर्फ यह कह रहे हो कि तापमान आरामदायक है तो इसे आरामदायक और असुविधाजनक में वर्गीकृत किया जा सकता है या यह केवल इत्मीनान सा है या ऐसा ही कुछ और कहा जा सकता है। लेकिन एक तरह से आपको लग सकता है कि यहाँ एक क्रम हो सकता है। लेकिन इस पल हमारे पास तापमान के सहज और असुविधाजनक के बीच के अंतर के लिए कोई क्रम नहीं है। यहाँ तापमान हमारी रुचि का एक वेरिएबल है और यहां आप देख सकते हैं कि तापमान एक नॉमिनल वेरिएबल है।

अब मान लीजिए कि तापमान फिर से हमारी रुचि का वेरिएबल है, लेकिन अब हम यहां यह जानने में दिलचस्पी रखते है कि कोई तरल कितना गर्म है- चाहे वह ठंडा हो, गुनगुना हो या गर्म हो। तो आप देखते हैं कि इस चर में एक क्रम है। हमे पता है कि गुनगुना, ठंडे से गर्म है, या गर्म गुनगुना से गर्म है, यहाँ वेरिएबल ओर्डिनल है। हालाँकि, यहाँ हमे नहीं पता कि गुनगुने और ठंडे पेय के बीच का अंतर गर्म और गुनगुने पेय के समान है या नहीं।

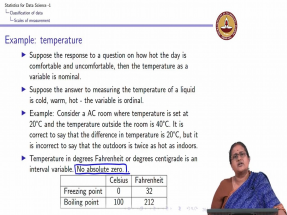
लेकिन अब मान लीजिए कि हम एक एसी कमरे (AC room) का तापमान माप रहे हैं जो बीस (20) डिग्री सेंटीग्रेड पर सेट है और उस कमरे के बाहर का तापमान चालीस (40) डिग्री सेंटीग्रेड है तो यह कहना सही होगा कि तापमान में अंतर बीस (20) डिग्री सेंटीग्रेड है, बिल्कुल ठीक।

मान लीजिए हमने एसी को 10 डिग्री या 14 डिग्री सेंटीग्रेड पर सेट किया और बाहर का तापमान 28 डिग्री हैं, तो मेरे लिए यह बताना बिल्कुल सही है कि 14 डिग्री सेंटीग्रेड का अंतर तापमान मे है। लेकिन मेरे लिए यह कहना बिल्कुल गलत है कि बाहर, घर के अंदर की तुलना में दोगुना गर्म है क्योंकि 40 डिग्री सेंटीग्रेड ,20 डिग्री सेंटीग्रेड से दोगुना गर्म नहीं होता है।

तो जब हम एक अंतराल पैमाने के बारे में बात करते हैं जैसे की हम अंतराल पैमाने में जानते है कि १० डिग्री सेंटीग्रेड और २० डिग्री सेंटीग्रेड के बीच का अंतर १० है, जो २० डिग्री सेंटीग्रेड और ३० डिग्री सेंटीग्रेड के बीच के अंतर के समान है, जो की 30 डिग्री सेंटीग्रेड और 40 डिग्री सेंटीग्रेड के बीच के अंतर के रूप में भी समान हैं।

लेकिन हम यह बयान नहीं दे सकते कि 40 डिग्री सेंटीग्रेड, 20 डिग्री सेंटीग्रेड से दोगुना गर्म होता है। यह कहना गलत होगा। तो हम यहाँ किन्हीं दो मूल्यों के बीच अंतर के बारे में बात कर सकते हैं, लेकिन यहां अनुपात का कोई मतलब नहीं है।

(स्लाइड समय देखें: 14:13)

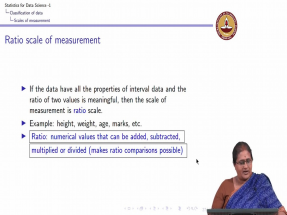


ऐसे ही जब हम तापमान के बारे मे और समझने की कोशिश करते हैं और जब हम सेल्सियस और फ़ारेनहाइट स्केल के बारे में बात करते हैं, तो इन मे कोई अब्सोलूट शून्य 0 नहीं होता है। सेल्सियस स्केल में क्रमश: शून्य (0) और सौ (100) को फ्रीजिंग पॉइंट (हिमांक) और बॉलिंग पॉइंट (क्वथनांक) के रूप में सेट किया जाता है, जबकि फ़ारेनहाइट में यह बत्तीस (32) और दो सौ बारह (212) होता है। केवल केल्विन (Kelvin) स्केल में आप शून्य (0) डिग्री पाएंगे जहां शून्य (0) का मतलब बिल्कुल तापमान नहीं है यानि अब्सोलूट ज़ीरो होता है। लेकिन जब हम सेल्सियस और फारेनहाइट स्केल के बारे में बात करते हैं तो हम समझते हैं कि इनमे कोई अब्सोलूट ज़ीरो (0) नहीं होता है।

तो जब हम इंटरवल स्केल के बारे में बात करते हैं तो हमारे लिए यह समझना अत्यंत महत्वपूर्ण है कि कोई अब्सोलूट ज़ीरो (0) नहीं है। हालाँकि इंटरवल स्केल और ओर्डिनल स्केल के बीच का अंतर यह है की इंटरवल स्केल में मानों के बीच का अंतर एक निश्चित माप की इकाई होती है, जबकि एक ओर्डिनल स्केल के अंतर के लिए माप की एक निश्चित इकाई होने की आवश्यकता नहीं है। जैसा की पिछले कथन मे अच्छे से बुरे के लिए तथा उत्कृष्ट से अच्छे के बीच मे समान अंतर नहीं है। यही इंटरवल स्केल और ओर्डिनल स्केल के बीच का मुख्य अंतर है।

माप का अंतिम पैमाना वह है जिसे हम माप के अनुपात पैमाने के रूप में संदर्भित करते हैं।

(स्लाइड समय देखें: 15:25)



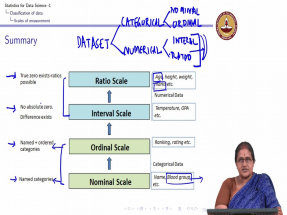
माप के अनुपात पैमाने (ratio scale) में इंटरवल डाटा के सभी गुण होते हैं और अनुपात एक बहुत ही सार्थक माप है। इस स्केल को अनुपात पैमाना यानि रैशियो स्केल कहा जाता है।उदाहरण के तौर पर ऊंचाई, वजन, अंक आदि। जैसा कि अगर एक व्यक्ति ने 300 अंक प्राप्त किए हैं तो अनुपात मे उसने दोगुना अंक प्राप्त किए हैं उस व्यक्ति के सापेक्ष जिसने 150 अंक प्राप्त किए हैं।

हम कह सकते हैं की 200 के स्कोर वाला व्यक्ति ,100 रन बनाने वाले व्यक्ति से दोगुना अच्छा है। हम यह भी कह सकते हैं कि एक व्यक्ति जिसने 100 विकेट लिए हैं, वह 50 विकेट लेने वाले व्यक्ति से दोगुना अच्छा है।

यहाँ हमारे पास अनुपात की धारणा हो सकती है जिसे हम यहां परिभाषित कर सकते हैं- वेरिएबल ऊंचाई, वजन, अंक, रन, विकेट ये सभी माप के अनुपात पैमाने के उदाहरण हैं। तो जब आपके पास एक वेरिएबल होता है जिसे अनुपात पैमाने में मापा जा सकता है तो आप इस पर सभी गणितीय या अंकगणितीय ऑपरेशन कर सकते हैं।

आप जोड़ सकते हैं, घटा सकते हैं, गुणा या भाग भी कर सकते हैं। लेकिन जब आप इंटरवल स्केल की बात करते हैं तो आप केवल अंतर के बारे में बात कर सकते हैं या आप केवल जोड़ और घटा सकते हैं। आपके पास कोई अब्सोलूट ज़ीरो (0) नहीं है, इसलिए अनुपातों के समकक्ष इसका कोई अर्थ नहीं है।

(स्लाइड समय देखें: 17:04)



जब भी आपको एक डाटा सेट प्रस्तुत किया जाता है तो यह पहचान कर लेने के बाद की वेरिएबल श्रेणीबद्ध हैं या संख्यात्मक हैं तो हमारे लिए यह समझना बेहद जरूरी है कि जब हमारे पास श्रेणीबद्ध डाटा होता है तो हमारे पास नॉमिनल (नाममात्र) और ओर्डिनल (क्रमिक) स्केल का उपयोग करना होता हैं। नॉमिनल (नाममात्र) के पैमाने के अंदर ,हमारे पास nominal है जिसे केवल श्रेणी का नाम दिया गया है, ओर्डिनल मे एक क्रम शामिल होता है। यहां क्रम के बीच का अंतर एक निश्चित अंतर नहीं है।

नाम, रक्त समूह एक श्रेणीबद्ध डाटा है या नॉमिनल स्केल के लिए उदाहरण है; ओर्डिनल (क्रमिक) पैमाने पर रैंकिंग, रेटिंग; एक क्रम है , लेकिन फिर रेटिंग में एक निश्चित क्रम होने की आवश्यकता नहीं है। इंटरवल स्केल के लिए अब्सोलूट ज़ीरो (0) नहीं है। यह संख्यात्मक डाटा या मात्रात्मक डाटा के लिए है। अब्सोलूट ज़ीरो (0) यानि निरपेक्ष शून्य (0) मौजूद नहीं है, लेकिन आनुपातिक अंतर (proportional difference) मौजूद है।

हम यहां अंतर का गणितीय संचालन कर सकते हैं। और अनुपात पैमाने में, अनुपात निकालना संभव हैं। आयु, ऊंचाई, वजन, अंक आदि सभी चर हैं जिन्हें अनुपात पैमाने पर मापा जा सकता है। तापमान, ग्रेड बिंदु, औसत, जैसे एक चार (4) GPA है जो दो (2) GPA से दोगुना अच्छा नहीं है, उन्हें इंटरवल स्केल पर मापा जा सकता है।

तो जब भी हमें डाटा दिया जाता है तो हम माप के पैमानों के बारे में क्यों रुचि रखते हैं? यहां आप देख सकते है की कोई अंकगणितीय संक्रिया संभव नहीं है। यहां हमे बस एक क्रम की कुछ उपस्थिति दिख सकती हैं। यहां हम जोड़, घटाव कर सकते हैं। यहां हम सभी अंकगणितीय ऑपरेशन कर सकते हैं। इसलिए जैसे ही हम वेरिएबल की पहचान कर लेते हैं हमारे द्वारा पूछे गए प्रश्नों का अर्थ स्पष्ट होने लगता है।

उदाहरण के लिए जब हमारे पास एक वेरिएबल है जो की एक ब्लड ग्रुप है तो हम यह सवाल नहीं पूछेंगे कि औसत ब्लड ग्रुप क्या है क्योंकि हम यहां इसमे किसी भी अंकगणितीय ऑपरेशन को परिभाषित नहीं कर सकते हैं। इसी तरह जब हम उम्र या अंक के बारे में बात करते हैं तो हम माप के इस पैमाने के आधार पर कुछ संख्यात्मक सारांशों के बारे में पूछ सकते हैं।

तो इस उदाहरण के साथ ही हम इस मॉड्यूल को समाप्त करते हैं। इस मॉड्यूल के अंत में आपको एक डाटा सेट को देखने में सक्षम होना चाहिए जो बहुत अच्छी तरह से व्यवस्थित है, आपको वेरिएबल को श्रेणीबद्ध या संख्यात्मक के रूप में पहचानें मे भी सक्षम होना चाहिए। और एक बार जब आप इन वेरिएबल की पहचान करने में सक्षम हो जाते हैं, तो आप आगे देखें कि माप के पैमाने क्या हैं, क्या यह नॉमिनल यानि नाममात्र है, क्या यह माप का एक ओर्डिनल यानि क्रमिक पैमाना है या फिर यह इंटरवल माप यानि अंतराल पैमाना है या माप का अनुपात पैमाना (ratio scale of measurement) है।

सांख्यिकी में लिखी गई अधिकांश पाठ्यपुस्तकों और पुस्तकों मे इन दोनों पैमानों (scales) को एक साथ मिला दिया गया; या तो अंतराल या अनुपात पैमाने के रूप में उल्लेख किया गया है। लेकिन फिर भी एक अंतर है और महत्वपूर्ण अंतर यह है कि अंतराल में कोई अब्सोलूट ज़ीरो (0) यानि निरपेक्ष शून्य (0) मौजूद नहीं है, जबकि अनुपात पैमाने में एक निरपेक्ष या एक वास्तविक शून्य (0) मौजूद है।

GLOSSARY

| ENGLISH WORD | HINDI WORD | MEANING |
| --- | --- | --- |
| Variable | वेरिएबल | चर |
| Arithmetic | अरिथमेटिक | अंकगणितीय |
| Operations | ऑपरेशन्स | संचालन |
| Absolute Zero | अब्सोलूट ज़ीरो | निरपेक्ष शून्य |
| Proportional Difference | प्रोपोर्शनल डिफरेंस | आनुपातिक अंतर |