Types of Correlation

(क) रेखीय सहसम्बंध (Linear Correlation)

(ख) अरेखीय सहसम्बंध (Non linear correlation)

(क) रेखीय सहसम्बंध :- जब दो चरों के बीच ऐसा सम्बंध है कि दोनों में समान अनुपात में वृद्धि या ह्रास हो तो इसे रेखीय सहसम्बंध कहा जाएगा।

(ख) अरेखीय सहसम्बंध :- वैसे वह रेखीय सहसम्बंध (curvilinear Correlation) और ही-2 वैषम्य सहसम्बंध (skew Correlation) भी कहा जाता है।

दो चरों के बीच सम्बंध की दिशा के आधार पर रेखीय सहसम्बंध (linear Correlation) के तीन प्रकार हैं :—

1. धनात्मक (Positive correlation) सहसम्बंध :-
धनात्मक सहसम्बंध दो चरों के बीच वह सहसम्बंध है, जहाँ एक चर में वृद्धि (increment) होने पर दूसरे चर में भी वृद्धि होती है, और एक चर में ह्रास (decrement) होने पर दूसरे चर में भी ह्रास हो तो इसे धनात्मक सहसम्बंध कहते हैं। जैसे :-
वृद्धि में वृद्धि होने पर उपलब्धि में भी वृद्धि होती है अतः इन दोनों चरों (वृद्धि एवं उपलब्धि) धनात्मक सहसम्बंध कहेंगे। (2) नींद का समय ↑ → थकान ↑
पढ़ाई का समय ↑ → परीक्षा बोर्ड ↑

2. ऋणात्मक सहसम्बंध (Negative Correlation) :-
दो चरों के बीच ऐसे सहसम्बंध ऋणात्मक सहसम्बंध कहते हैं, जहाँ एक चर में वृद्धि होने पर दूसरे चर में ह्रास होता है और एक चर में ह्रास होने पर दूसरे चर में वृद्धि हो।

उदाहरण :- थकान के बढ़ने पर उत्पादन घटता है और तथा थकान के घटने पर उत्पादन बढ़ता है। अतः इन दोनों चरों (थकान एवं उत्पादन) के बीच ऐसे सहसम्बंध को ऋणात्मक सहसम्बंध कहा जाएगा।

3. Zero correlation (शून्य सहसम्बंध) :→

जब दो चरों के बीच कोई सम्बंध नहीं होता है।

शून्य सहसम्बंध का अर्थ यह है कि दो चरों के बीच कोई रेखीय सम्बंध (linear relationship) नहीं है। एक चर में वृद्धि अथवा ह्रास होने पर दूसरे चर में न तो वृद्धि ही और न ह्रास हो तो दोनों चरों के बीच ऐसे सहसम्बंध को शून्य सहसम्बंध कहा जाएगा। इसे शून्य द्वारा व्यक्त किया जाता है।

इस प्रकार स्पष्ट है कि सहसम्बंध के तीन प्रकार हैं — धनात्मक, ऋणात्मक तथा शून्य। किन्तु शून्य सहसम्बंध को सहसम्बंध कहना ही अनुचित है। इसका कारण यह है कि, सहसम्बंध की परिभाषा के अनुसार दो चरों के बीच धनात्मक या ऋणात्मक सम्बंध होना आवश्यक है। अतः वास्तविक अर्थ में सहसम्बंध के केवल दो प्रकार हैं, जिन्हें धनात्मक सहसम्बंध तथा ऋणात्मक सहसम्बंध कहते हैं।

— o —

Correlation मापने के तरीके (Method of measuring correlation) :→

1. Pearson's Product Moment Correlation (r)
   या Product moment correlation method
   - ⓐ यह सबसे लोकप्रिय method है।
   - ⓑ Continuous data के लिए उपयुक्त विधि है।
   - ⓒ इसका मान −1 से +1 के बीच रहता है।
   - ⓓ +1 = Perfect positive, −1 = Perfect Negative, 0 = no correlation
   - ⓔ The Rank Difference Correlation method

2. Spearman's Rank order (ρ)
   - ⓐ यह Rank data के लिए प्रयोग होता है।
   - ⓑ यह small sample और ordinal data के लिए उपयोगी विधि है। ⓒ जब actual scores नहीं हो लेकिन ranks हों

Formula - $\rho = 1 - \frac{6\Sigma D^2}{N(N^2-1)}$

(e) The Biserial correlation method (f) The Point Biserial correlation method (g) The Tetrachoric correlation method (h) The Phi - coefficient Method (i) The coefficient of Contingency (j) The Partial Correlation method (k) Multiple correlation Method (कई तरीके होते हैं)

3. Scatter Diagram Method :-

(a) यह visual method है। (b) यह तरीके आसान और graph में points plot करके relationship का अंदाजा लगाया जाता है। (c) X और Y axis पर values plot करके relationship देखा जाता है। (d) इसमें dots के बीच रेखा में होती है। इसमें slope correlation के प्रकार को आसानी से समझने के लिए।

प्रकार	विवरण	उदाहरण
1. Positive Correlation (धनात्मक सहसंबंध)	एक चर के बढ़ने पर दूसरा चर भी बढ़ता है।	पढ़ाई का समय ↑ परीक्षा का स्कोर ↑, अपराध में उम्र
2. Negative Correlation (ऋणात्मक सहसंबंध)	एक variable के बढ़ने से दूसरा चर घटता है।	mobile का उपयोग ↑ → नींद की गुणवत्ता ↓
Zero Correlation (शून्य सहसंबंध)	variable में कोई सहसंबंध नहीं होता है।	व्यक्ति की चप्पल का आकार और उसकी बुद्धिमत्ता
Perfect Correlation (+1 / -1)	100% संबंध (बहुत अच्छा pure)	शुद्ध वार्षिक संबंध $y = c$ या x
Partial Correlation	दो variables के बीच संबंध लेकिन अन्य variable का प्रभाव हटा दें	study time और marks का संबंध जब intelligence को constant रखा जाए।
Multiple Correlation	एक dependent variable का संबंध कई independent variable से	marks का संबंध study time + Attendance + IQ से

Correlation Coefficient (r)

Correlation की ताकत आपने के लिए Correlation coefficient निकाला जाता है। यह हमेशा -1 से +1 के बीच होता है।

r का मान	Interpretation
+1.00	Perfect Positive correlation
+0.70 से +0.99	High Positive correlation
+0.30 से +0.69	Moderate Positive correlation
0.00 से +0.29	Low Positive correlation
0.00	No correlation
-0.30 से -0.69	Moderate Negative correlation
-0.70 से -0.99	High Negative correlation
-1.00	Perfect Negative correlation

Correlation Method के उपयोग :-

Prediction :- अगर हम height और weight का correlation जानते हैं, तो height देखकर weight का अनुमान लगाया जा सकता है।

Selection & classification :- इसमें Aptitude test का अंक (marks) से correlation देखकर इन सही test का चुनाव किया करते हैं।

Tools validation verification :→ किसी test की validity जानने के लिए अन्य established test से correlation निकाला जाता है।

① Planning Research :→ चरों (variables) के सम्बंध जानकर आगे के प्रयोग (experiment) की design बनाई जाती है।

Limitation (सीमाएँ) याद रखने के

② Causation नहीं बताता :→ Strong correlation का मतलब यह नहीं कि एक variable दूसरे variable का कारण है।

③ Spurious correlation :→ जहाँ 2 से variables में relationship केवल संयोग (Concidental) होता है।

जैसे :- Ice cream sales ↑ और Drowing case ↑
(असल कारण - गर्मी का मौसम)

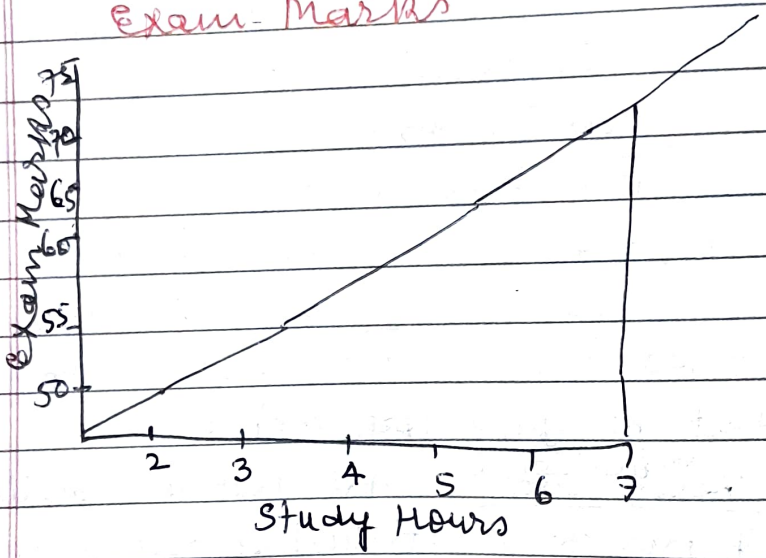
classmate
Date
Page

④ Measurement Error :→ अगर data गलत है तो गलत correlation आ सकता है।

⑤ Outliers का प्रभाव :→ कुछ extreme values पूरी calculation बिगाड़ सकता है।

Pearson's का एक छोटा Numerical example के द्वारा (scatter diagram से इसके concept को और clear करते हैं।

Scatter diagram : Study Hours Vs Exam Marks



इस scatter diagram में study hours और exam marks का strong positive correlation दिख रहा है।

Pearson's $r = 0.996 \ (\approx 1)$

⇒ इसका मतलब है कि पढ़ाई के समय और परीक्षा में प्राप्त अंकों के बीच बहुत strong correlation है। अर्थात जैसे-2 study hours बढ़ते हैं, वैसे-2 exam marks भी बढ़ते हैं।

Frequency Distribution of Data & Graphic Presentations.

जब हम किसी data set को व्यवस्थित (organize) करके इस तरह प्रस्तुत करते हैं कि वह व्यवस्थित

Spearman's Rank Correlation table

Spearman's ρ (rho) = 1.0 (Perfect Positive Correlation)

क्योंकि study hours और exam marks दोनों perfectly ordered ranks में हैं (जैसे-2 study Hours बढ़ते हैं, marks भी बढ़ते हैं) इसलिए Spearman's ρ बिल्कुल +1 आया है।

Comparison table of Pearson's r और Spearman's ρ की विशेषताएँ (differences, use, cases, formula side by side को समझते हैं ताकि याद रखना आसान हो जाए।

Pearson's r Vs Spearman's ρ का comparison

पहलू	Pearson's r	Spearman's ρ (rho)
परिभाषा	दो continuous variables के बीच linear relationship और direction को मापता है।	दो variables के ranks के बीच relationship को मापता है।
Data का प्रकार	Interval और Ratio scale (continuous data)	ordinal data (rank data)
Formula	$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}$	$\rho = 1 - \frac{6\sum D^2}{N(N^2-1)}$
relationship	केवल linear relationship को मापता है।	linear और Non-linear दोनों monotonic relationship को माप सकता है।
Outliers का प्रभाव	बहुत ज्यादा effect करते	कम effect करता है (क्योंकि rank का use करता है)
Value का range	-1 से +1	-1 से +1
Perfect correlation का मतलब	सब points बिल्कुल straight line में हों	जब ranks बिल्कुल same order में हों।
Use case	जब data continuous और normally distributed हो।	जब data ordinal हो, या linearity doubtful हो।

Summary :-

⇒ Pearson's r → सबसे अच्छा continuous normally distributed data के लिए बहुत ही उपयोगी विधि है।

⇒ Spearman's ρ → जब Ranks हों या data linear न हो या Outliers हों तब इसका उपयोग अधिक फायदेमंद होता है।

Spearman's formula :-

$$\rho = 1 - \frac{6\Sigma D^2}{N(N^2-1)}$$

$$\text{Pearson's } r = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\Sigma(X-\bar{X})^2(Y-\bar{Y})^2}}$$