Q. No. बहुसहसंबंध (Multicollinearity) की व्याख्या

प्रतिगमन विश्लेषण (Regression Analysis) में अनेक स्वतंत्र चरों (Independent Variables) का उपयोग करके आश्रित चर (Dependent Variable) पर उनके प्रभाव का अध्ययन किया जाता है। प्रतिगमन विश्लेषण का एक महत्वपूर्ण आधार यह है कि स्वतंत्र चर एक-दूसरे से अत्यधिक संबंधित न हों। जब दो या दो से अधिक स्वतंत्र चर आपस में बहुत अधिक सहसंबंध (High Correlation) रखते हैं, तो इस स्थिति को बहुसहसंबंध (Multicollinearity) कहा जाता है।

बहुसहसंबंध प्रतिगमन विश्लेषण की एक गंभीर समस्या है क्योंकि इससे यह निर्धारित करना कठिन हो जाता है कि आश्रित चर पर किस स्वतंत्र चर का वास्तविक प्रभाव कितना है। इस कारण प्रतिगमन गुणांकों (Regression Coefficients) का सही अनुमान लगाना कठिन हो जाता है और शोध के परिणामों की विश्वसनीयता कम हो सकती है।

बहुसहसंबंध की परिभाषा (Definition)

बहुसहसंबंध (Multicollinearity) वह स्थिति है जिसमें प्रतिगमन मॉडल के दो या दो से अधिक स्वतंत्र चर (Independent Variables) आपस में अत्यधिक सहसंबंधित (Highly Correlated) होते हैं। अर्थात वे एक-दूसरे के बारे में लगभग समान जानकारी प्रदान करते हैं, जिसके कारण उनके व्यक्तिगत प्रभाव का सही आकलन करना कठिन हो जाता है।

सरल शब्दों में–

जब स्वतंत्र चर एक-दूसरे से इतने अधिक जुड़े हों कि उनमें से किसी एक के प्रभाव को अलग-अलग मापना कठिन हो जाए, तब उसे बहुसहसंबंध कहते हैं।

उदाहरण (Example)

मान लीजिए किसी शोधकर्ता को छात्रों के परीक्षा परिणाम (Marks) पर निम्नलिखित कारकों का प्रभाव जानना है–

प्रतिदिन पढ़ाई के घंटे

प्रति सप्ताह पढ़ाई के घंटे

प्रति माह पढ़ाई के घंटे

ये तीनों चर आपस में अत्यधिक संबंधित हैं क्योंकि यदि कोई छात्र प्रतिदिन अधिक पढ़ता है, तो उसके साप्ताहिक और मासिक अध्ययन घंटे भी अधिक होंगे। ऐसी स्थिति में यह बताना कठिन हो जाएगा कि परीक्षा परिणाम पर किस चर का प्रभाव अधिक है। यही बहुसहसंबंध कहलाता है।

एक अन्य उदाहरण में यदि किसी व्यक्ति की आय (Income) और उसके मासिक व्यय (Expenditure) दोनों को किसी मॉडल में शामिल किया जाए, तो दोनों में अक्सर उच्च सहसंबंध पाया जाता है। इससे भी बहुसहसंबंध उत्पन्न हो सकता है।

बहुसहसंबंध के प्रकार (Types of Multicollinearity)

1. पूर्ण बहुसहसंबंध (Perfect Multicollinearity)

जब एक स्वतंत्र चर दूसरे स्वतंत्र चर का पूर्ण (100%) रैखिक संयोजन हो, तब इसे पूर्ण बहुसहसंबंध कहते हैं।

उदाहरण:

यदि $X_2 = 2X_1$ हो, तो दोनों चरों में पूर्ण बहुसहसंबंध होगा।

2. अपूर्ण बहुसहसंबंध (Imperfect Multicollinearity)

जब स्वतंत्र चरों के बीच उच्च लेकिन पूर्ण (100%) सहसंबंध न हो, तब इसे अपूर्ण बहुसहसंबंध कहते हैं। वास्तविक जीवन में अधिकांश शोधों में यही स्थिति पाई जाती है।

बहुसहसंबंध के कारण (Causes of Multicollinearity)

- स्वतंत्र चरों के बीच अत्यधिक सहसंबंध होना।
- समान प्रकार के आँकड़ों का उपयोग करना।
- छोटे प्रतिदर्श (Small Sample Size) का होना।
- ऐसे चर शामिल करना जो एक-दूसरे पर निर्भर हों।
- समय-श्रृंखला (Time Series) आँकड़ों में समान प्रवृत्ति (Trend) होना।
- गलत मॉडल विनिर्देशन (Model Specification)।
- अत्यधिक संख्या में स्वतंत्र चरों को शामिल करना।
- बहुसहसंबंध के दुष्प्रभाव (Consequences of Multicollinearity)

1. प्रतिगमन गुणांक अस्थिर हो जाते हैं।

स्वतंत्र चरों में थोड़ा-सा परिवर्तन भी गुणांकों के मान में बड़ा परिवर्तन ला सकता है।

2. मानक त्रुटि (Standard Error) बढ़ जाती है।

गुणांकों का अनुमान कम सटीक हो जाता है।

3. t-परीक्षण के परिणाम अविश्वसनीय हो जाते हैं।

महत्वपूर्ण चर भी महत्वहीन दिखाई दे सकते हैं।

4. स्वतंत्र चरों के वास्तविक प्रभाव की व्याख्या कठिन हो जाती है।

5. नीति-निर्धारण एवं निर्णय लेने में कठिनाई होती है।

6. मॉडल की विश्वसनीयता कम हो जाती है।

7. पूर्वानुमान (Prediction) की गुणवत्ता प्रभावित हो सकती है।

बहुसहसंबंध की पहचान (Detection of Multicollinearity)

सहसंबंध मैट्रिक्स (Correlation Matrix) द्वारा।

Variance Inflation Factor (VIF) द्वारा।

Tolerance Test द्वारा।

Eigen Value एवं Condition Index द्वारा।

प्रतिगमन परिणामों का विश्लेषण करके।

बहुसहसंबंध को दूर करने के उपाय (Remedies of Multicollinearity)

अत्यधिक सहसंबंध वाले स्वतंत्र चरों में से किसी एक को हटा देना।

अधिक एवं बेहतर गुणवत्ता वाले आँकड़े एकत्र करना।

संबंधित चरों को मिलाकर नया चर बनाना।

Principal Component Analysis (PCA) का उपयोग करना।

Ridge Regression जैसी आधुनिक तकनीकों का प्रयोग करना।

मॉडल का पुनः विनिर्देशन (Respecification) करना।

अनावश्यक स्वतंत्र चरों को हटाना।

बहुसहसंबंध के लाभ (Advantages)

यद्यपि बहुसहसंबंध सामान्यतः समस्या मानी जाती है, फिर भी कुछ परिस्थितियों में इसके कुछ लाभ हो सकते हैं–

यदि उद्देश्य केवल पूर्वानुमान (Prediction) करना हो, तो इसका प्रभाव अपेक्षाकृत कम हो सकता है।

कई संबंधित चरों के बीच संबंधों को समझने में सहायता मिल सकती है।

वास्तविक जीवन के आर्थिक एवं सामाजिक आँकड़ों की प्रकृति को समझने में मदद मिलती है।

निष्कर्ष (Conclusion)

बहुसहसंबंध (Multicollinearity) प्रतिगमन विश्लेषण की एक महत्वपूर्ण समस्या है, जिसमें स्वतंत्र चर आपस में अत्यधिक संबंधित होते हैं। इसके कारण प्रतिगमन गुणांकों का सही अनुमान लगाना कठिन हो जाता है तथा परिणामों की विश्वसनीयता प्रभावित होती है। इसलिए किसी भी शोध में प्रतिगमन विश्लेषण करने से पहले बहुसहसंबंध की जाँच करना आवश्यक है। यदि यह समस्या पाई जाए, तो उपयुक्त सांख्यिकीय तकनीकों का उपयोग करके इसे कम या समाप्त किया जाना चाहिए। इस प्रकार शोध के निष्कर्ष अधिक सटीक, विश्वसनीय तथा उपयोगी बनते हैं।