SIMPLE AND LINEAR REGRESSION ANALYSIS

BY

DR. SANTOSH KUMAR LAL

DEPT. OF COMMERCE

SARIYA COLLEGE, SURIYA

SIMPLE AND LINEAR REGRESSION ANALYSIS

Regression analysis is a statistical technique used to study the relationship between a **dependent variable (Y)** and one or more **independent variables (X)**. It helps us understand how changes in an independent variable are associated with changes in the dependent variable and can also be used for prediction.

सरल एवं रैखिक प्रतिगमन विश्लेषण

प्रतिगमन विश्लेषण (Regression Analysis) एक सांख्यिकीय तकनीक है जिसका उपयोग **आश्रित चर** (Dependent Variable – Y) और एक या अधिक **स्वतंत्र चरों** (Independent Variable – X) के बीच संबंध का अध्ययन करने के लिए किया जाता है। इसके द्वारा यह समझने में सहायता मिलती है कि स्वतंत्र चर में परिवर्तन होने पर आश्रित चर में किस प्रकार परिवर्तन हो सकता है तथा भविष्यवाणी (Prediction) भी की जा सकती है।

मुख्य उद्देश्य — MAIN OBJECTIVES

1. Variables के बीच संबंध को मापना।
2. Dependent variable के मूल्य का अनुमान लगाना।
3. Independent variable के प्रभाव का अध्ययन करना।
4. Business, economics, science और social sciences में forecasting करना।
5. यह समझना कि X में परिवर्तन Y से किस प्रकार संबंधित है।

MEANING — अर्थ

Simple regression analysis examines the relationship between **one dependent variable and one independent variable**. Simple Regression Analysis में **एक आश्रित चर** (Y) और **एक स्वतंत्र चर** (X) के बीच संबंध का अध्ययन किया जाता है।

उदाहरण:

- Advertising expenditure → Sales
- Study hours → Examination marks
- Income → Consumption
- Temperature → Ice-cream sales

यदि हम यह जानना चाहते हैं कि **advertising expenditure बढ़ने पर sales में कितना परिवर्तन हो सकता है**, तो Simple Regression का प्रयोग किया जा सकता है।

SIMPLE LINEAR REGRESSION — सरल रैखिक प्रतिगमन

- जब X और Y के बीच संबंध को एक straight line द्वारा व्यक्त किया जाता है, तो उसे Simple Linear Regression कहते हैं।

Regression Equation

- Y=a+bX

जहाँ:

- **Y** = Dependent variable / आश्रित चर
- **X** = Independent variable / स्वतंत्र चर
- **a** = Intercept / Constant
- **b** = Regression coefficient / Slope
- Interpretation

यदि:

- Y=20+5X
- तो इसका अर्थ है:
- जब X = 0 हो, तो predicted Y = 20।
- X में 1 unit की वृद्धि होने पर Y में औसतन 5 units **की वृद्धि** होगी।

REGRESSION LINE — प्रतिगमन रेखा

A regression line is the line that represents the estimated relationship between X and Y. In simple linear regression, it is generally written as:

Y^=a+bX

Here Y^ represents the **predicted value of Y**.

Regression Line वह रेखा है जो X और Y के बीच अनुमानित संबंध को प्रदर्शित करती है। Simple Linear Regression में इसे सामान्यतः इस प्रकार लिखा जाता है:

Y^=a+bX

यहाँ Y^, Y के **अनुमानित** (predicted) **मूल्य** को दर्शाता है।

SLOPE AND INTERCEPT – ढाल एवं अंतःखंड

A. Intercept (a)

The intercept is the expected value of Y when X is zero.

Intercept उस Y के अनुमानित मूल्य को दर्शाता है जब X का मूल्य शून्य हो।

B. Slope (b)

The slope indicates the expected change in Y for a one-unit change in X.

Slope बताता है कि X में **एक** unit **परिवर्तन** होने पर Y में औसतन कितना परिवर्तन अपेक्षित है।

Slope के प्रकार

- **b > 0:** Positive relationship
- **b < 0:** Negative relationship
- **b = 0:** No linear relationship according to the fitted slope

उदाहरण:

- $Y=50+3X$
- यहाँ $b=3$ है। अतः X में 1 unit की वृद्धि होने पर Y में औसतन 3 units की वृद्धि का अनुमान है।

LEAST SQUARES METHOD — न्यूनतम वर्ग विधि

- Simple Linear Regression में regression line के coefficients सामान्यतः **Least Squares Method** से estimate किए जाते हैं।
- इसका उद्देश्य ऐसी line प्राप्त करना है जिससे actual और predicted values के बीच errors के squared values का कुल योग न्यूनतम हो।

जहाँ:

- Y = Actual value
- Y^ = Predicted value
- Y—Y^ = Residual/Error

सरल भाषा मे
Least Squares Method ऐसी regression line खोजती है जिसके लिए actual observations और fitted values के बीच deviations के वर्गों का योग सबसे कम हो।

The least squares method selects the line that minimizes the sum of squared residuals between observed and predicted values.

RESIDUAL OR ERROR — अवशेष/त्रुटि

Regression में:

$e = Y - \hat{Y}$

जहाँ e residual है।

Residual actual value और predicted value के बीच अंतर है।

A residual is the difference between the observed value and the value predicted by the regression model.

उदाहरण:

- यदि actual sales = ₹100 और predicted sales = ₹90,
- $e = 100 - 90 = 10$
- अर्थात residual = 10।

ASSUMPTIONS OF LINEAR REGRESSION — प्रमुख मान्यताएँ

Linear regression के उचित उपयोग के लिए कुछ महत्वपूर्ण assumptions होती हैं:

1. Linearity — **रैखिकता**

- X और Y के बीच संबंध को उचित रूप से linear form में represent किया जा सके।

2. Independence — **स्वतंत्रता**

- Observations/residuals में अनावश्यक dependence नहीं होनी चाहिए।

3. Homoscedasticity — **समान** variance

- Residuals की variability predictor के विभिन्न levels पर लगभग समान होनी चाहिए।

4. Normality of errors — **त्रुटियों की सामान्यता**

- Classical statistical inference में residual/errors की normality महत्वपूर्ण हो सकती है, विशेषकर छोटे samples में।

5. Absence of serious outliers

- बहुत प्रभावशाली outliers regression results को काफी प्रभावित कर सकते हैं।

6. For multiple regression: limited multicollinearity

- Independent variables के बीच अत्यधिक correlation नहीं होना चाहिए।

SIMPLE REGRESSION के APPLICATIONS – उपयोग

Simple Linear Regression का प्रयोग विभिन्न क्षेत्रों में किया जाता है:

- **Business**

Advertising expenditure के आधार पर sales forecasting।

- **Economics**

Income के आधार पर consumption का अनुमान।

- **Education**

Study hours और marks के बीच relationship का analysis।

- **Agriculture**

Rainfall के आधार पर crop yield का estimation।

- **Marketing**

Price changes के आधार पर demand का prediction।

- **Finance**

कुछ financial variables के आधार पर returns या अन्य outcomes का statistical modelling।

ADVANTAGES — लाभ

1. Simple to understand and interpret.
2. Useful for prediction and forecasting.
3. Quantifies the relationship between variables.
4. Can support business and scientific decision-making.
5. Provides useful statistical measures such as coefficients and R2.

1. समझना और interpret करना अपेक्षाकृत आसान है।
2. Prediction और forecasting में उपयोगी है।
3. Variables के संबंध को numerical form में व्यक्त करता है।
4. Business और scientific decision-making में सहायता करता है।
5. Regression coefficients और R2 जैसे महत्वपूर्ण statistical measures प्रदान करता है।

LIMITATIONS — सीमाएँ

1. Regression relationship को model करता है, लेकिन causation **अपने-आप सिद्ध नहीं करता**।
2. Outliers results को प्रभावित कर सकते हैं।
3. गलत model specification से misleading conclusions मिल सकते हैं।
4. Predictions data range से बहुत बाहर करने पर unreliable हो सकती हैं।
5. Correlated observations या violated assumptions statistical inference को प्रभावित कर सकती हैं।
6. केवल R2 देखकर model की quality का निर्णय नहीं करना चाहिए।

QUICK REVISION — त्वरित पुनरावृत्ति

Simple Regression: 1 dependent + 1 independent variable

Simple Linear Regression: Y^=a+bX

a: Intercept

b: Slope

Residual: e=Y−Y^

Estimation method: Least Squares

- **R2:** Explained variation का proportion
- **Positive b:** X बढ़ने पर predicted Y बढ़ता है
- **Negative b:** X बढ़ने पर predicted Y घटता है
- **Multiple Regression:** 1 dependent + 2 या अधिक independent variables
- **Key caution:** Regression ≠ automatic proof of causation