

Programming for Big Data

Saeed Iqbal Khattak

Centre for Healthcare Modelling & Informatics
Faculty of Information Technology,
University of Central Punjab, Lahore

April 27, 2021



Outline

- ▶ Correlation vs. Regression
- ▶ Simple Linear Regression Model
- ▶ Introduction to Regression Analysis
- ▶ Types of Relationships
- ▶ Regression Model

Correlation vs. Regression

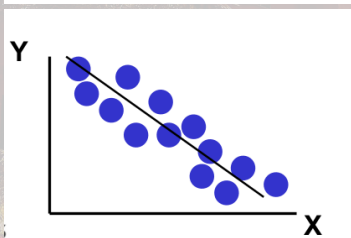
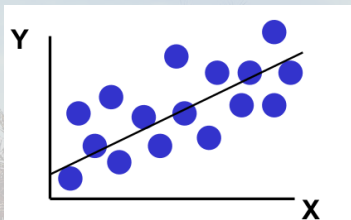
- ▶ A scatter diagram can be used to show the relationship between two variables
- ▶ Correlation analysis is used to measure strength of the association (linear relationship) between two variables
- ▶ Correlation is only concerned with strength of the relationship.

Regression Analysis

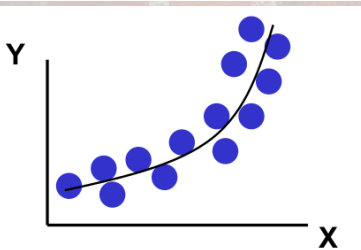
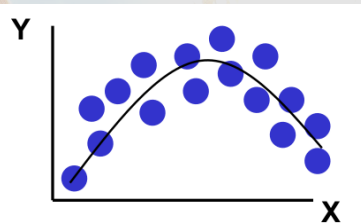
- ▶ Regression analysis is used to:
 - ▶ Predict the value of a dependent variable based on the value of at least one independent variable.
 - ▶ Explain the impact of changes in an independent variable on the dependent variable
- ▶ **Dependent variable:** the variable we wish to predict or explain.
- ▶ **Independent variable:** the variable used to explain the dependent variable

Types of Relationships

Linear Relationships

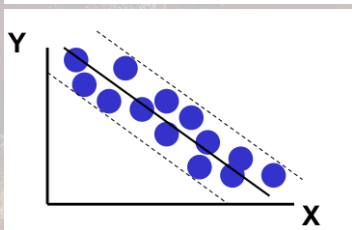
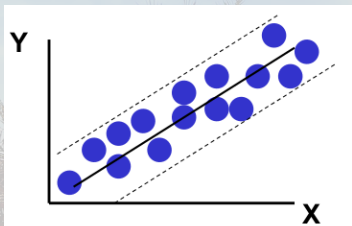


CurviLinear Relationships

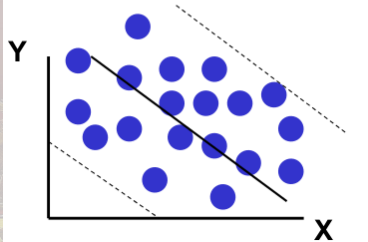
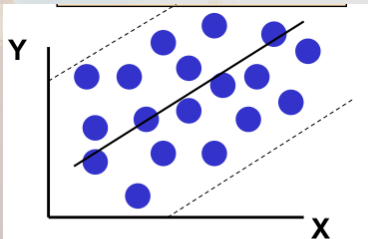


Types of Relationships

Strong Relationships



Weak Relationships



Regression Model

- ▶ Only **one independent variable, X**
- ▶ Relationship between X and Y is described by a linear function.
- ▶ Changes in Y are assumed to be caused by changes in X.

The diagram shows the equation $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ with the following labels and arrows:

- Dependent Variable** points to Y_i .
- Population Y intercept** points to β_0 .
- Population Slope Coefficient** points to β_1 .
- Independent Variable** points to X_i .
- Random Error term** points to ϵ_i .

Below the equation, two blue brackets group the terms:

- A bracket under $\beta_0 + \beta_1 X_i$ is labeled **Linear component**.
- A bracket under ϵ_i is labeled **Random Error component**.

Figure: Simple Linear Regression Model

Simple Linear Regression Model

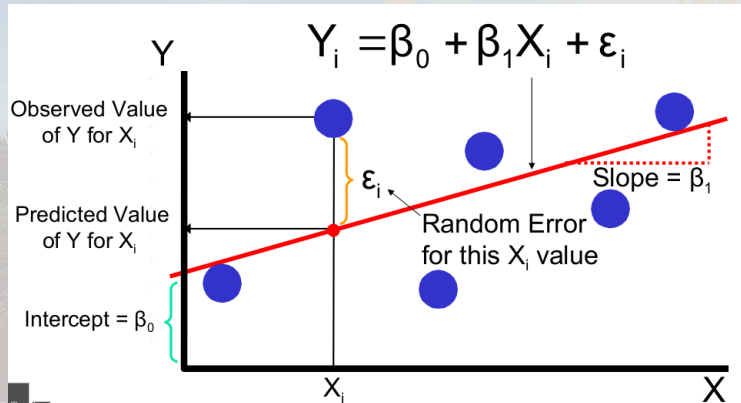


Figure:

Python Code – Jupyter Notebook

Thank You