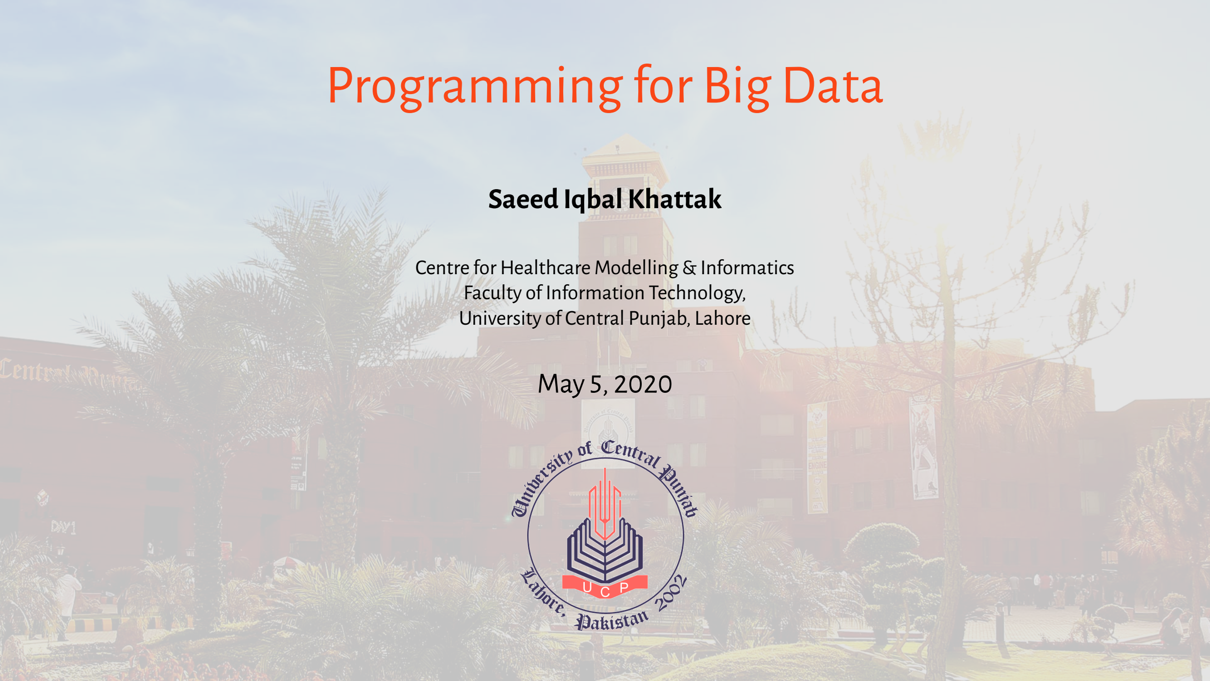


Programming for Big Data

Saeed Iqbal Khattak

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

May 5, 2020

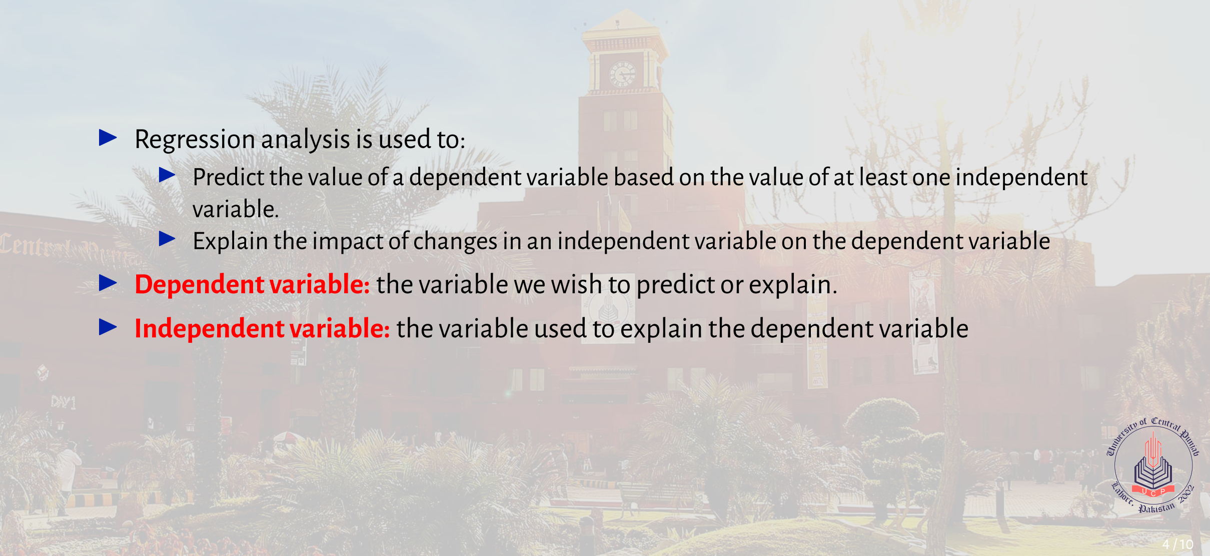
Outline

- 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.

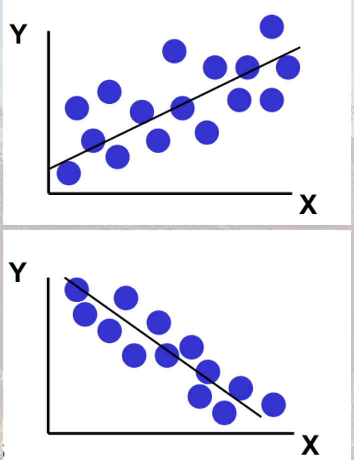
Regression Analysis



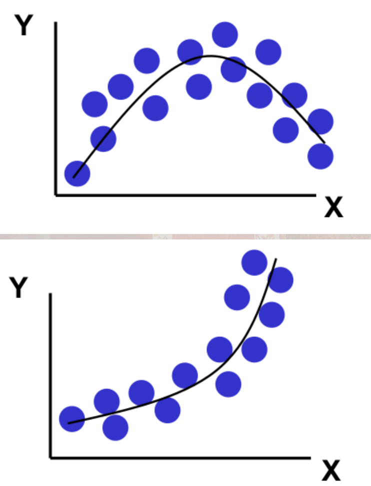
- Regression analysis is used to:
  - Predict the value of a dependent variable based on the value of at least one independent variable.
  - Explain the impact of changes in an independent variable on the dependent variable
- **Dependent variable:** the variable we wish to predict or explain.
- **Independent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

Types of Relationships

Linear Relationships

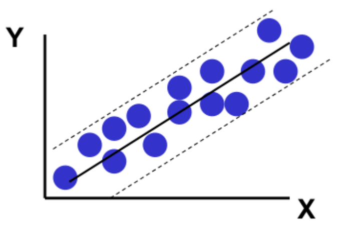


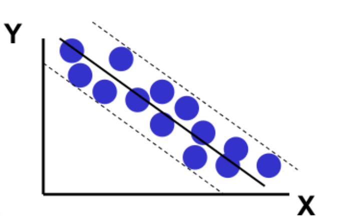
CurviLinear Relationships



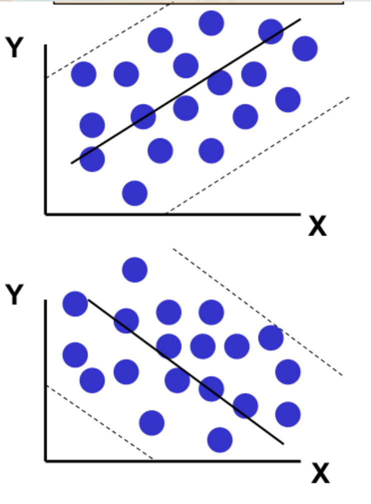
Types of Relationships

Strong Relationships



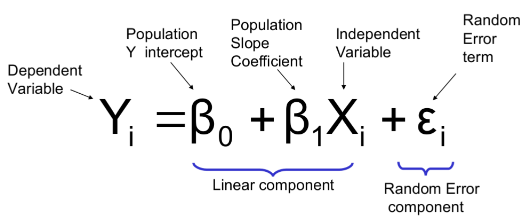


Weak Relationships



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.



$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Figure: Simple Linear Regression Model

Simple Linear Regression Model

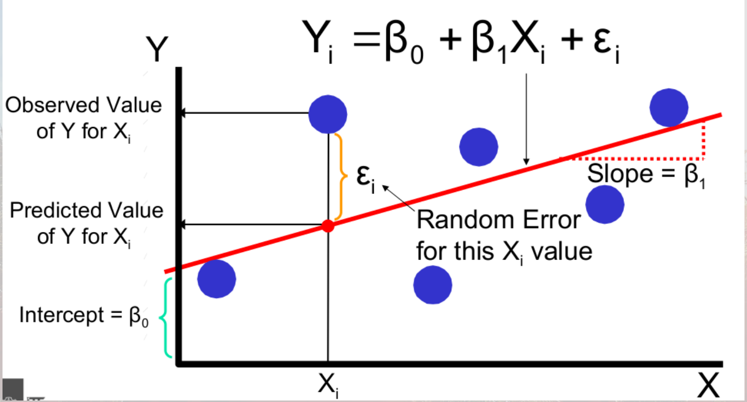
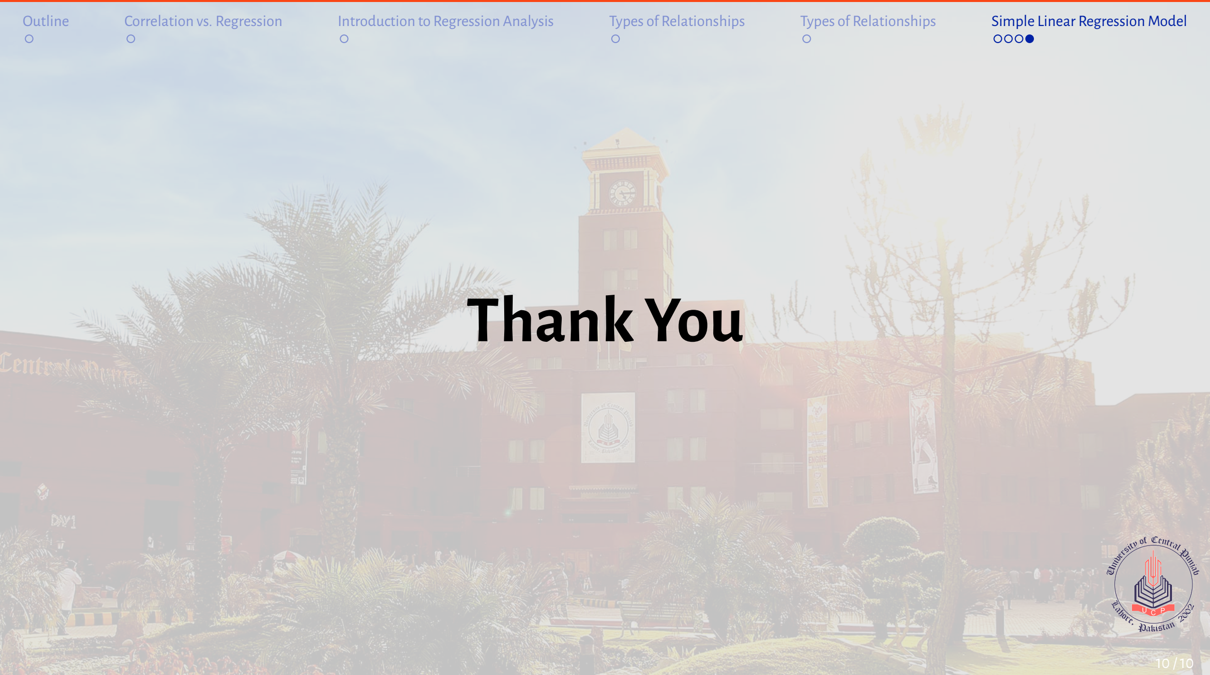

$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
Y
Observed Value of Y for X_i
Predicted Value of Y for X_i
Intercept = β_0
ε_i
Random Error for this X_i value
Slope = β_1
X_i
X

Figure:

Python Code – Jupyter Notebook



Thank You