

# Linear Regression Program Documentation

## Objective

The objective of this program is to implement simple linear regression to model the relationship between two variables,  $x$  and  $y$ , using the least squares method. The program computes the regression line:

$$y = a + bx$$

where  $a$  is the intercept,  $b$  is the slope, and then predicts the value of  $y$  for a given input  $X$ .

## Concept

**Linear Regression** is a statistical technique used to model and analyze the relationship between a dependent variable ( $y$ ) and an independent variable ( $x$ ).

- Slope ( $b$ ): Measures how much  $y$  changes for a unit change in  $x$ .
- Intercept ( $a$ ): Value of  $y$  when  $x = 0$ .

The least squares method minimizes the sum of squared differences between observed values and predicted values.

The formulas used:

$$b = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2}$$
$$a = \frac{\sum y - b\sum x}{n}$$

## Program Features

- Reads input data from a file (input.txt) and writes output to a file (output.txt).
- Handles multiple test cases.
- Computes regression coefficients  $a$ (intercept) and  $b$ (slope).

- Predicts  $y$  for a given input  $X$ .
- Prints the regression line equation and predicted value to both console and file.
- Detects and reports if regression cannot be computed (e.g., division by zero in slope calculation).

## Input Format

- File: input.txt
- Each test case starts with a line containing the number of data points  $n$ .
- Next  $n$  lines contain pairs of  $x$   $y$  values separated by space.
- After the  $n$  data points, a single value  $X$  is given for prediction.

### Example:

5

1 2

2 3

3 5

4 4

5 6

10

Explanation:

- 5 data points:  $(1,2)$ ,  $(2,3)$ ,  $(3,5)$ ,  $(4,4)$ ,  $(5,6)$
- Predict  $Y$  for  $X = 10$

Multiple test cases can be included sequentially in the same file.

## Output Format

- File: output.txt (also printed on console)
- For each test case:

Test Case 1

$y = a + bx$

Predicted Y = \_\_ for X = \_\_

- If regression cannot be computed:

Test Case 1

Error: Regression cannot be computed.

#### **Example Output:**

Test Case 1

$y = 1.3 + 0.9x$

Predicted Y = 10.3 for X = 10

## Algorithm

1. Read number of points (n) from input.
  2. Read n data points (x and y values).
  3. Read the input value X for which prediction is required.
  4. Compute regression coefficients:
    - o  $\text{sumX} = \Sigma x$
    - o  $\text{sumY} = \Sigma y$
    - o  $\text{sumXY} = \Sigma(x * y)$
    - o  $\text{sumX2} = \Sigma(x^2)$
    - o Slope:  $b = (n * \text{sumXY} - \text{sumX} * \text{sumY}) / (n * \text{sumX2} - \text{sumX}^2)$
    - o Intercept:  $a = (\text{sumY} - b * \text{sumX}) / n$
  5. Predict Y:  $Y = a + b * X$ .
  6. Print and write output (regression line and predicted Y).
  7. Repeat for all test cases until end of input file.
-

## Features of This Program

- Handles multiple test cases from the same input file.
- Computes linear regression coefficients automatically.
- Predicts Y for any given X.
- Outputs are written to file and printed on console.
- Detects errors where regression cannot be computed (e.g., all x-values are identical).
- Easy to extend for other types of regression or additional input validation.