

Product Requirements Document: Simulation Analysis Tool

1. Introduction

This document outlines the product requirements for the Simulation Analysis Tool, a Python script that simulates a dataset based on a linear equation, performs linear regression, and visualizes the results. The tool is designed to help users understand the relationship between a dependent and an independent variable, the effect of random noise on a linear relationship, and the use of linear regression to model data.

2. Goals and Objectives

- To provide a tool for understanding the relationship between a dependent and an independent variable.
- To visualize the effect of random noise on a linear relationship.
- To demonstrate the use of linear regression to model data.

3. Features and Functionality

3.1. Data Generation

- Generates a specified number of data samples.
- Uses a configurable linear equation ($y = ax + b + r$) to generate data.
- Allows for customization of the equation parameters (a , b) and the amount of random noise.

3.2. Linear Regression

- Calculates the line of best fit for the generated data using the least squares method.

3.3. Visualization

- Creates a scatter plot of the generated data.
- Overlays the calculated regression line on the scatter plot.
- Displays the data generation and regression equations in a text box on the plot.

3.4. Output

- Saves the generated plot as a PNG image.
- Saves the equations to a text file.

4. User Persona

Data analysts, students, and researchers who need to understand and visualize linear regression.

5. Technical Requirements

- **Programming Language:** Python 3
- **Libraries:** NumPy, Matplotlib

6. Assumptions and Dependencies

- The user has a Python 3 environment with NumPy and Matplotlib installed.
- The script assumes that the relationship between the variables is linear.

7. Future Enhancements

- Allow the user to input the equation parameters and the number of samples through command-line arguments.
- Support other types of regression (e.g., polynomial regression).
- Provide more detailed statistical analysis of the results (e.g., R-squared value).
- Create a graphical user interface (GUI) for interacting with the simulation.