

# Unraveling Linear Relationships: Regression Analysis with SEMMA

Nick Kornienko

September 21, 2023

## Abstract

This paper delves into a detailed regression analysis using the SEMMA methodology on a dataset to understand the linear relationship between the independent variable  $x$  and the dependent variable  $y$ . The focus is also on addressing the nuances and intricacies of the dataset, including handling missing values to ensure reliable model training and predictions.

## 1 Introduction

Understanding relationships between variables is a cornerstone in data science. This paper focuses on exploring and modeling the linear association between variables  $x$  and  $y$  using the structured SEMMA methodology.

## 2 Exploration and Modification

The exploration phase unveiled the dataset's structure and the presence of missing values in the target variable  $y$ . These were addressed by removing rows with missing target values, ensuring the robustness and reliability of the subsequent modeling phase.

## 3 Modeling and Assessment

A Linear Regression model was employed and optimized to predict the target variable  $y$ . The assessment was conducted using the Mean Squared Error, offering insights into the model's reliability and the average squared difference between the actual and predicted values.

## 4 Conclusion

The structured approach of the SEMMA methodology allowed for comprehensive exploration and analysis, addressing the dataset's specific nuances and en-

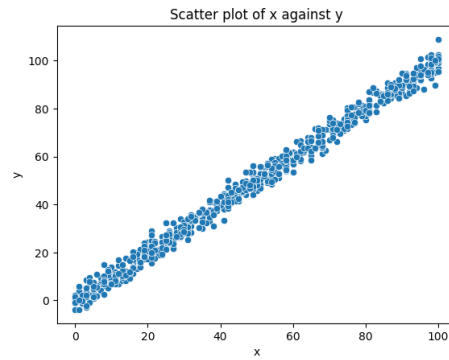


Figure 1: Scatter plot of  $x$  against  $y$

suring the model was well-understood, optimized, and assessed. This methodology proved pivotal in uncovering underlying patterns and relationships in the dataset, leading to reliable and insightful predictions.