

Linear Regression for Model Overfitting

Our study dives into the connection between two important variables, x and y . We've collected a solid dataset of 1500 observations for each variable and want to use a simple linear regression to see if y can predict x .

Here's how we did it step by step in Excel:

1. **Setting Up:** We started by putting our data into Excel. Then, we went to the "Data" tab and clicked on "Data Analysis" to open the Analysis ToolPak.
2. **Defining Ranges:** We picked the range for y as our dependent variable and the range for x as our independent variable. Since we labeled our data, we made sure to check the "Labels" box.
3. **Creating Results Sheet:** We created a new sheet to display our results.
4. **Options Setup:** In the Analysis ToolPak, we selected all the options for residuals to get a thorough look at our results.
5. **Running Regression:** After setting everything up, we clicked "Ok" to run the regression.

On our results sheet, we got a bunch of useful information:

Summary Output: This table gave us key stats like the correlation between x and y , how much of the variation in x can be explained by y (coefficient of determination), and more.

ANOVA Table: Here, we found the p -value, which tells us if our regression model is statistically significant.

Residual Outputs: These tables and plots helped us understand the difference between our predicted x values and the actual ones.