Linear Regression:

In statistical modeling, regression analysis is a set of statistical processes for estimating the relationships between a dependent variable (often called the outcome or response variable, or a label in machine learning parlance) and one or more error-free independent variables (often called regressors, predictors, covariates, explanatory variables or features).

OR

Regression analysis is a way of mathematically sorting out which of those variables does indeed have an impact. It answers the questions: Which factors matter most? Which can we ignore? How do those factors interact with one another? And, perhaps most important, how certain are we about all these factors?

In regression analysis, those factors are called “variables.” You have your dependent variable — the main factor that you’re trying to understand or predict.. And then you have your independent variables— the factors you suspect have an impact on your dependent variable.

Let’s look at an example from one of the papers I’m working on.

We wanted to see the answer to the following question

When considering regulations for AI, which of the following statements best aligns with your perspective?

1. I prefer a system where clear, detailed rules outline what is required to ensure compliance and prevent misuse of AI.
2. I prefer a system that emphasizes broad principles such as ethical use and transparency, allowing for flexibility in how these are achieved.

We have so many other variables

* Income level
* Education level
* Source of Information on AI

Plot the Data

Draw a line through it

A graph with a red line

AI-generated content may be incorrect.

Income = 14,867.37 + 5,317.66 x Education + error term

*A graph with red line

AI-generated content may be incorrect.*

From our fitted logistic regression model, the equation is:

*A black and white math equation

AI-generated content may be incorrect.*

where:

intercept = -6.8924

0.000105 is the coefficient for income meaning income positively influences the likelihood of choosing B.

Higher income increases P(B), meaning people with higher incomes are more likely to choose B.

A regression line always has an error term because, in real life, independent variables are never perfect predictors of the dependent variables. Rather the line is an estimate based on the available data. So, the error term tells you how certain you can be about the formula. The larger it is, the less certain the regression line.

Typically you start a regression analysis wanting to understand the impact of several independent variables. You might include not just income but also education level and source of information about AI