

Linear Regression Analysis

Regression analysis is a technique used to determine the association between two random variables X and Y. Linear regression attempts to explain this relationship with a straight line fit to the data. The linear regression model assumes that

$$Y = a + bX + e$$

where the "residual" e is a random variable with mean zero. The coefficients a and b are determined by the condition that the sum of the square residuals is as small as possible.

Some questions that linear regression might address are:

- Could the variables x and y realistically be independent (in the population), or can we conclude that there is an association between them?
- If the variables are associated, how strong is the association?
- How does the outcome for the response variable depend on the value of the explanatory variable, and which observations are unusual?