

The Method of Least Squares

When we fit a regression line to a set of points, we assume that there is some unknown linear relationship between Y and X , and that for every one-unit increase in X , Y increases by some set amount on average. Our fitted regression line enables us to predict the response, Y , for a given value of X . But for any specific observation, the actual value of Y can deviate from the predicted value. The deviations between the actual and predicted values are called errors, or residuals. The better the line fits the data, the smaller the residuals (on average).

How do we find the line that best fits the data? In other words, how do we determine values of the intercept and slope for our regression line? Intuitively, if we were to manually fit a line to our data, we would try to find a line that minimizes the model errors, overall. But, when we fit a line through data, some of the errors will be positive and some will be negative. In other words, some of the actual values will be larger than their predicted value (they will fall above the line, and some of the actual values will be less than their predicted values (they'll fall below the line).

If we add up all of the errors, the sum will be zero. So how do we measure overall error? We use a little trick. We square the errors and find a line that minimizes the sum of the squared errors. This method, the method of least squares, finds values of the intercept and slope coefficients that minimize the sum of the squared errors.

Statistical Thinking for Industrial Problem Solving

Copyright © 2020 SAS Institute Inc., Cary, NC, USA. All rights reserved.

Close