

6.4: Effect Size

In the last section we concerned ourselves with testing the hypothesis that the dependent variable did indeed depend upon the hypothesized independent variable or variables. It may be that we find an independent variable that has some effect on the dependent variable, but it may not be the only one, and it may not even be the most important one. Remember that the error term was placed in the model to capture the effects of any missing independent variables. It follows that the error term may be used to give a measure of the "goodness of fit" of the equation taken as a whole in explaining the variation of the dependent variable, Y .

The effect size is given by the formula:

$$R^2 = \frac{SS_{\text{Reg}}}{SS_{\text{Total}}}$$

where SS_{Reg} (or SSR) is the regression sum of squares, the squared deviation of the predicted value of y from the mean value of y ($\hat{y} - \bar{y}$), and SS_{Total} (or SST) is the total sum of squares. Figure 13.9 shows how the total deviation of the dependent variable, y , is partitioned into these two pieces.

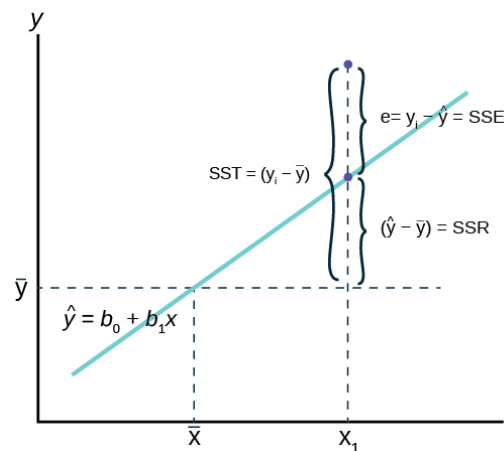


Figure 13.9

Figure 13.9 shows the estimated regression line and a single observation, x_1 . Regression analysis tries to explain the variation of the data about the mean value of the dependent variable, y . The question is, why do the observations of y vary from the average level of y ? The value of y at observation x_1 varies from the mean of y by the difference $(y_i - \bar{y})$. The sum of these differences squared is SST , the sum of squares total. The actual value of y at x_1 deviates from the estimated value, \hat{y} , by the difference between the estimated value and the actual value, $(y_i - \hat{y})$. We recall that this is the error term, e , and the sum of these errors is SSE , sum of squared errors. The deviation of the predicted value of y , \hat{y} , from the mean value of y is $(\hat{y} - \bar{y})$ and is the SS_{Reg} , sum of squares regression. It is called "regression" because it is the deviation explained by the regression. (Sometimes the SS_{Reg} is called SS_{Model} for sum of squares model because it measures the change from using the mean value of the dependent variable to using the model, the line of best fit. In other words, it measures the deviation of the model from the mean value of the dependent variable, y , as shown in the graph.)

Because $SS_{\text{Total}} = SS_{\text{Reg}} + SS_{\text{Error}}$, we see that the effect size is the percent of the variance, or deviation in y from its mean value, that is explained by the equation when taken as a whole. R^2 will vary between 0 and 1, with 0 indicating that none of the variation in y was explained by the equation and a value of 1 indicating that 100% of the variation in y was explained by the equation.

While a high R^2 is desirable, remember that it is the tests of the hypothesis concerning the existence of a relationship between a set of independent variables and a particular dependent variable that was the motivating factor in using the regression model. The goal of choosing the regression analysis is to validate a statistical relationship developed by some theory. Increasing the number of independent variables will have the effect of increasing R^2 . But the goal is not to add as many independent variables as you possibly can; instead it is to select robust independent variables as informed by theories and/or empirical literature.

This page titled 6.4: Effect Size is shared under a [CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/) license and was authored, remixed, and/or curated by [Yang Lydia Yang](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

- **13.4: The Regression Equation** by OpenStax is licensed CC BY 4.0. Original source: <https://openstax.org/details/books/introductory-business-statistics>.