

6.5: Multicollinearity

Multicollinearity

Our discussion earlier indicated that like all statistical models, the OLS regression model has important assumptions attached. Each assumption, if violated, has an effect on the ability of the model to provide useful and meaningful estimates. Here we will look at the effects on OLS estimates if the independent variables are correlated. We take up multicollinearity because it is so often prevalent in social sciences studies and it often leads to frustrating results.

The OLS model assumes that all the independent variables are independent of each other. This assumption is easy to test for a particular sample of data with simple correlation coefficients. Correlation, like much in statistics, is a matter of degree: a little is not good, and a lot is terrible.

The goal of the regression technique is to tease out the independent impacts of each of a set of independent variables on some hypothesized dependent variable. If two independent variables are interrelated, that is, correlated, then we cannot isolate the effects on Y of one from the other. In an extreme case where x_1 is a linear combination of x_2 , correlation equal to one, both variables move in identical ways with Y . In this case it is impossible to determine the variable that is the true cause of the effect on Y .

The correlation has the same effect on the regression coefficients of both these two variables. In essence, each variable is “taking” part of the effect on Y that should be attributed to the collinear variable. This results in biased estimates.

Furthermore, multicollinearity often results in failing to reject the null hypothesis that the X variable has no impact on Y when in fact X does have a statistically significant impact on Y . Said another way, the large standard errors of the estimated coefficient created by multicollinearity suggest statistical insignificance even when the hypothesized relationship is strong.

This page titled [6.5: Multicollinearity](#) is shared under a [CC 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>.