

CHAPTER OVERVIEW

5: Predicting Responses

PREDICTION is typically the primary goal of most regression modeling projects. That is, the model developer wants to use the model to estimate or predict the system's response if it were operated with input values that were never actually available in any of the measured systems. For instance, we might want to use the model we developed using the Int2000 data set to predict the performance of a new processor with a clock frequency, a cache size, or some other parameter combination that does not exist in the data set. By inserting this new combination of parameter values into the model, we can compute the new processor's expected performance when executing that benchmark program.

Because the model was developed using measured data, the coefficient values necessarily are only estimates. Consequently, any predictions we make with the model are also only estimates. The `summary()` function produces useful statistics about the regression model's quality, such as the R^2 and adjusted R^2 values. These statistics offer insights into how well the model explains variation in the data. The best indicator of any regression model's quality, however, is how well it predicts output values. The R environment provides some powerful functions that help us predict new values from a given model and evaluate the quality of these predictions.

[5.1: Data Splitting for Training and Testing](#)

[5.2: Training and Testing](#)

[5.3: Predicting Across Data Sets](#)

[5.4: Section 5-](#)

[5.5: Section 6-](#)

This page titled [5: Predicting Responses](#) is shared under a [CC BY-NC 4.0](#) license and was authored, remixed, and/or curated by [David Lilja](#) ([University of Minnesota Libraries Publishing](#)) via [source content](#) that was edited to the style and standards of the LibreTexts platform.