Print

Scenario

Imagine you're a real estate agent and you want to know which houses are the best ones to spend your time trying to sell. Which houses should you focus on selling to make the greatest commission? Is it the house with the biggest yard or the most bathrooms? Does it matter how many fireplaces are in the house or whether the garage is attached or detached?

We know that multiple regression will tell us the relationship between a predictor and the response while adjusting for all other predictors in the model. So we can compute the association between basement area and sale price while adjusting for the number of bathrooms and other predictors in the model.

But which predictors do we include in the model in the first place?

Deciding on the best model for your needs might seem overwhelming. For example, if you have 10 predictors, there are more than 1000 possible models. However, you can use several techniques for model selection to help you produce a manageable number of candidate models. Then you can use your subject-matter expertise and knowledge of the question to choose the best model for your purpose.

In this section, we'll produce candidate models using stepwise model selection techniques, including forward, backward, and stepwise selection.

Close