Association vs. Prediction

L. J. Wei

Harvard University

Abstract

We were told as students that there were two aims for doing regression analysis. The first one is to quantify the association between the response and "independent" variables. The second one is to make prediction of response based on those independent variables. Since a model is just an approximation to the true one (if there is one), the association parameters may vary from model to model (especially in the presence of confounders). Moreover, it seems difficult (if not impossible) to validate association even with an independent data set. On the other hand, prediction may be validated even with a working model. We provide a couple of examples to show how to utilize prediction idea for model selections. We also propose inference procedures under working models.