4.9 Cautionary notes

Before proceeding, a few cautionary notes are appropriate with respect to building models. First, the procedures of forward selection, backward elimination and step-wise selection are in no way guaranteed to find the best overall model. Second, even if the search for the “best” overall model is done, the result will depend on which selection criterion is used (adjusted *R*^{2}, *AIC*, *AIC*$_c$ or *BIC*). Third, DISTLM fits a linear combination of the **X** variables, which may or may not be appropriate in a given situation (e.g., see the section on Linkage trees in chapter 11 of Clarke & Gorley (2006)). As a consequence, it is certainly always appropriate to spend some time with the **X** variables doing some diagnostic plots and checking out their distributions and relationships with one another as a preliminary step. Fourth, the particular predictor variables that are chosen in a model should not be interpreted as being necessarily causative[85]. The variables chosen may be acting as proxies for some other important variables that either were not measured or were omitted from the model for reasons of parsimony. Finally, it is *not* appropriate to use a model selection procedure and then to take the resulting model and test for its significance in explaining variability. This approach uses circular and therefore invalid logic, because it is already the purposeful job of the model selection procedure to select useful explanatory variables. To create a valid test, the inherent bias associated with model selection would need to be taken into account by performing the selection procedure anew with each permutation. Such a test would require a great deal of computational time and is not currently available in DISTLM[86]. In sum, model-building using the DISTLM tool should generally be viewed as an *exploratory hypothesis-generating* activity, rather than a definitive method for finding the one “true” model.

---

[85] *Unless* a predictor variable has been expressly manipulated experimentally in a structured and controlled way to allow causative inferences, that is.

[86] The approach of including the selection procedure as part of the test is, however, available for examining non-parametric relationships between resemblance matrices as part of PRIMER’s BEST routine. See pp. 124-125 in chapter 11 of Clarke & Gorley (2006) for details.

Revision #2
Created 12 August 2022 11:11:44 by Arden
Updated 5 September 2022 13:13:09 by Arden