Table of Contents

Ordination analysis

Forward selection (constrained ordination)

Theory R functions Examples Exercise 🔒

Forward selection is a procedure for selecting a subset of explanatory variable from the set of all variables available for constrained ordination (RDA or CCA). The goal is to reduce the number of explanatory variables entering the analysis, while keeping the variation explained by them to maximum. Suitable mostly in case of observational studies, where many (often highly intercorrelated) environmental variables are recorded, to reduce their number (and to simplify the story); not useful for experimental studies with balanced design of treatment application.

The simplified sequence of steps is the following:

1. first, test the significance of the global test with all explanatory variables included; if it is significant, you may proceed to forward selection, while if it is not, it is better not to (remember that even with randomly generated explanatory variables you have rather good chance to select some of them as significant during forward selection).
2. use each variable one by one as explanatory in constrained ordination, and record the explained variation (this variation represents simple (or marginal) effect of each variable);
3. sort variables according to variation explained by them with the highest values at the top;
4. check whether the variation explained by the best variable is significant using Monte Carlo permutation test - if yes, include it to the model, if not, stop the selection;
5. use each of remaining explanatory variables and check how much variation they (each separately) explain if put as explanatory (with the already selected variable acting as covariable);
6. sort again the variables according to the decreasing variation explained by them (now this variation represents partial effect of this variable) and choose the one explaining the most; test whether the variation is significant, and if yes, select it into the model; if not, stop the selection;
7. continue by step 5 until the variation explained by the best variable is not significant.

The simplest method is *forward selection*, which is adding explanatory variables one by one; backward selection, in contrary, starts from the full model and deletes variables which the least decreases the total explained variation. Combination of both approaches is forward-backward selection, in which in every step the analysis checks whether some of already included variables cannot be removed to improve the model.

Significance of the variables is one of the possible stopping rules (once the best variable is not significant, the selection is stopped). Alternative stopping rule is reaching the adjusted R^2 of the global model (Blanchet et al. 2008): first calculate adjusted variation explained by all explanatory variables (global model); if during the forward selection the adjusted variation explained by selected variables reaches the R^2_{adj} of the global model, the selection will be stopped (available in function `ordiR2step` in `vegan` and `forward.sel` in `packfor`).

From:
https://anadat-r.davidzeleny.net/ - **Analysis of community ecology data in R**

Permanent link:
https://anadat-r.davidzeleny.net/doku.php/en:forward_sel?rev=1549787026

Last update: **2019/02/10 16:23**