Biometry Lab 2012

Logistic Regression

A colleague provides you with data (“H\_sulphuria.xls”) containing presence/absence records of a predaceous diving beetle (Dytiscidae: *Heterosternuta sulphuria*) that is of conservation concern in Arkansas. The data also includes environmental data that was collected from a GIS database at various spatial scales: watershed, riparian, and local.

You have been asked to compare a few specific hypotheses, listed below. Please check that the assumptions are met for each logistic regression model. Be sure to check for collinearity among predictor variables if multiple predictors are used. In the results tab of the Binary Logistic Regression module in Systat, be sure to save residuals which will contain several diagnostic metrics. Use AIC or BIC when making comparisons among models.

Report any models and/or predictor variables that significantly affect occurrence of *H. sulphuria*.

This species is known to be a headwater specialist that is sensitive to watershed area (‘SHED\_AREA’), so this should be included as a predictor in all models.

H1) Percent urbanization in watersheds (‘SHED\_URBAN’) significantly affects the probability of *H. sulphuria* occupying stream sites.

H2) Percent urbanization in watersheds (‘SHED\_URBAN’) and percent forest in riparian zones (‘RIPARIAN\_FOREST’) together provide better estimates—compared to H1—of the probability of *H. sulphuria* occupying sites.

H3) Percent forest in riparian zones (‘RIPARIAN\_FOREST’) significantly interacts with urbanization in watersheds (‘SHED\_URBAN’). In other words, the negative effect of urbanization in the watershed is dependent on how much of the riparian zone is forested.

After evaluating and comparing each of these hypotheses, please use a stepwise procedure to identify other predictor variables that may be important. Again, be sure to check for collinearity among predictor variables. What are the potential problems with stepwise selection procedures?

The American burying beetle (Silphidae: *Nicrophorus americanus*)is the only federally endangered insect in Arkansas. It must locate and bury small carcasses which it buries in underground brood chambers to provision larvae. This reproductive requirement limits *N. americanus* to habitats with soils suitable for digging and vegetation communities that support adequate avian and mammal populations to provide carrion. There are two prominent competing hypotheses about what type of vegetation communities are preferred, forests vs. grasslands. The data set ‘N\_americanus*.*xls’ contains presence/absence data for *N. americanus* and other guild members along with environmental data collected from a GIS database from within a one-mile radius surrounding each sample site.

Please evaluate and compare the following hypotheses using logistic regression. Check that the assumptions are met for each logistic regression model. Be sure to check for collinearity among predictor variables. In the results tab of the Binary Logistic Regression module in Systat, be sure to save residuals which will contain several diagnostic metrics. Report any models and/or predictor variables that significantly affect occurrence of *N. americanus*. Use AIC or BIC when making comparisons among models.

H1) *N. americanus* occurrence is determined by the proportion of forest within a one-mile radius (‘FOREST’).

H2) *N. americanus* occurrence is determined by the proportion of grassland within a one-mile radius (‘GRASSLAND’).

H3) *N. americanus* occurrence is determined by the dominant soil type within a one-mile radius (‘SOIL\_DOM’).

H4) *N. americanus* occurrence is related to the occurrence of other carrion beetles, particularly ‘Deltachilum’, a large dung beetle, and ‘tomentosus’, a small ubiquitous burying beetle.

H5) Land cover (either ‘FOREST’ or ‘GRASSLAND’) and dominant soil (‘SOIL\_DOM’) must both be considered to predict occurrence of *N. americanus.*\*\* Be sure to identify ‘SOIL\_DOM’ as a categorical variable.

H6) Land cover, dominant soil, and biotic interactions provide the best model when considered all together (*e.g.* ‘GRASSLAND’, ‘SOIL\_DOM’, ‘Deltachilum’)