Canonical Correspondence Analysis

KEY

Thu Nov 10 11:10:14 2022

## Variation of bird species composition at the San Dieguito River Park estuary

This data set is interesting both because it is from a local source, and because we can learn a method for the proper way to incorporate circular variables, such as month, into your analysis.

To begin, import the data:

Make a list of the species (all columns are species except the first four):

Load the vegan library:

Run a correspondence analysis using the cca() function with only the bird data (no “environmental” matrix yet):

Get a report of variation associated with each axis, and the running total across axes:

#### Question: what proportion of variance is explained by the first two CA axes?

Get a screeplot of the eigenvalues:

#### Question: is there a noticeable dropoff in the size of the eigenvalues across these first eight CC axes? Between which axes?

Construct the biplot:

#### Question: according to the biplot, which species are most strongly associated with CA1? Which species would you expect to see in large numbers in the points at CA1 values of 2 or more?

#### Question: which species are most strongly associated with CA2? Which species would you expect to see in large numbers in points with CA2 scores of -3 or less?

Calculate the CA 1 scores:

Correlate the bird counts with these CA 1 scores:

Plot buffleheads, as an example of a strong positive correlation:

#### Question: Is the relationship between BUFF and CA1 linear? Does the correlation coefficient accurately represent the relationship between BUFF counts and CA1 scores?

Now plot elegant terns, as an example of a strong negative correlation:

#### Question: can you see why there is a negative correlation between ELTE and CA1? Is the relationship linear?

Plot brown pelicans as an example of a correlation near 0:

#### Question: there are two reasons why you might get a low correlation between counts for a species and an axis: species that are restricted to a narrow range in the middle of the axis, or species that are found in roughly equal numbers along the entire axis. Which is true here for brown pelicans?

## Canonical correspondence analysis

We will use month and year as our “environmental matrix”, but we want month to be represented by springness and winterness to avoid having December and January at opposite ends of the scale.

First, we need an ordered factor for month:

Convert the months to radians on a circle:

Calculate springness and add it to the birds data set:

Calculate winterness and add it to the birds data set:

Now run the CCA, relating the birds to year, springness, and winterness:

#### Question: the eigenvalue for CCA1 is 0.30925. Does this indicate that it explain 30.9% of the variation in bird species, or does it explain 30.9% of the 24.5% that is explained across all of the CCA axes?

Make a triplot, with site scores, species scores, and vectors indicating the predictors:

#### Question: which season has the least variation from year to year in species composition? Hint: look for the spiders with the shortest legs, and see wheich quadrant of the graph they’re in.

#### Question: which species have the highest relative frequencies in each season?

Get a randomization test for the CCA to see if the amount of variation in bird community composition that is explained by season and time is statistically significant:

#### Question: which of the predictors are statistically significant predictors of species composition?

To calculate loadings, first extract the site scores:

Then correlate the bird counts by the site scores:

#### Question: which species are most strongly positively correlated with CCA1? Which are strongly negatively correlated with CCA1?

Correlate the site scores with the predictor variables:

#### Question: which axis is most strongly associated with change over time in species composition (year)?

To better understand the change over time, correlate the species that have the highest loadings on CCA3 with year:

#### Question: based on these correlations, how are the four species changing over time (which are increasing, which are decreasing)?

To make sure that we’re justified in using year as a linear predictor in the model, fit a cca with year as a linear predictor, followed by year as a factor:

#### Question: is year as a factor significant, after the linear change over time is accounted for? What does this mean - is there variation from year to year that isn’t a directional trend over time?