Instructions for using Species Richness program

Note: The program will do 10 000 iterations, outputting the species richness from each random iteration into a text file (space-delimited). Each row of the output file will contain the calculated species richness for each sample and for each defined group, where a group is 1 or more samples combined and species richness is calculated across the set of samples in the group (e.g. a species present in only one of two samples within a group would increase the species richness for that group).

Currently the program is limited to reading in abundances for 200 species and a total abundance (number of individuals summed across all samples and all species) of 50 000. The maximum number of samples for randomization is 200, organized into no more than 20 groups. (These maximums are easily changed; please contact me if you need them increased, or want them decreased to reduce program run time.)

During testing, a set of 4 samples in 2 groups of 2 each, comprised of 166 species and 22 221 individuals required approximately 15 minutes to complete on a Pentium 4 2.8 GHz desktop and 30 minutes on a Pentium 4 2.0 GHz laptop.

While running, the program will print to the screen iteration numbers as it proceeds, so you will be able to assess how far along it is (not to mention whether or not it is actually making progress!). The numbers will show up in blocks at a time; more than 2 minutes without a new set of numbers showing up on the screen means that something is wrong (use CTRL-C to exit the program).

TO RUN THE RANDOMIZATION PROGRAM:

Preparing the data input file

The program reads data from a plain text file.

There are three parts to the file: samples, groups, species. Each of these labels should be written in the text file, each on its own line, and in that order. On the line following the label 'samples', the number of samples included in the dataset (up to 200) should be given, followed by the abundance of each sample, each on its own line. Here is how it will look:

samples
4
6290
3494
4293
6144

These samples may in turn be categorized into groups, e.g. 2 northern and 2 southern sample sites. If you your samples are in groups, include the label "groups" in your input file and indicate on the next line the number of groups into which the samples fall. Follow this with the number of samples in each group. For example,

groups
2
2
2

Note that when you enter sample abundances you need to have samples from the same group together. You also need to indicate the number of samples in each group based on the order in which you input sample abundances. Groups may contain only one sample and the groups may be arranged in any order, as long as all the samples for a group are entered together when you give abundances (i.e. you could have the first sample be one group, the next 3 samples be a second group, and the last 2 samples be a third group).

If your samples are not in groups, simply leave the "groups" label out of the file.

The final section of the file, after the label 'species' will contain species' labels and abundances arranged in a single column, with the abundance of a species (for all samples combined) on the line beneath that species label. For example,

species
G.alpha
32
G.beta
59
G.gamma
300

etc.

The easiest way to generate this file is to create a file in your database with 2 columns – species identifier and total abundance, and then save that file in comma-delimited format. The number of samples and sample abundances can be placed in the top row and the number of groups and number of samples in each group can be placed in the second row. You can than open the comma-delimited file in a word processor and do a find & replace to replace all commas with a line-return. Save the result as a plain text file. **Note: this must be a DOS plain text file with line endings set to "CR".** The finished text file should have no blank lines.

Species names/identifiers must not contain spaces and must be no more than 30 characters long. Be sure there is no return after the last number in the file. For an example of a correctly formatted data file, see "Grixti_data.txt" on the website containing these instructions and the program file (http://www.jeanrichardson.ca/spp_richness).

The data text file should be placed in the same folder as the program file. If the text file is not located in the same folder as the randomization program, you will need to enter the complete path name of the file when you give the program the file name. Note that clicking and dragging the file name does not work

To run the program, simply double-click on its name.

At any point during the programme execution (e.g. if you give it the wrong file name, or enter something wrong) you can exit the program by typing CTRL-C.

You will first be asked for the name of the file in which the data are contained. (If the file does not exist, the program will abort).

1. Next you will be asked to give a name to the file into which results are to be put. This file does not need to exist (and will be overwritten if it does exist). Again, dragging from windows will not work.

2. The program will tell you how many species and the total abundance of all species that it has read in from the file.

3. Let it run! It may take a while, but it will let you know what iteration it is on as it goes through them. As long as it puts numbers on the screen once and awhile, it is doing fine. The window that popped up when you started the program will automatically close when it is done. You can then find the results file and open it in your favourite spreadsheet software to find the expected mean species richness per sample and per group, as well as the expected distribution if there is no effect of sample.

   e.g. In Excel, you can find the critical values for species richness for a 2-tailed test by using the NORMINV function to calculate the value of species richness above and below which 2.5 % of the distribution falls. First, calculate the mean for the distribution (the column of numbers generated by the program for the appropriate sample) using the function =AVERAGE() in an empty cell on the sheet. Second, calculate the standard deviation of the same distribution using the function =STDEV() in another cell on the sheet. Then use the function =NORMINV(0.025, average, stdev) to get critical value for the left-hand tail. The function =NORMINV(0.975, average, stdev) will give you the critical value for the right-hand tail. Any observed value falling above or below these critical values suggest a result significantly ($p<0.05$) different from the null expectation .