**Homework – Diversity Indices Practice**

**Bio 580 – Dr. Whitenack**

**Due Thurs Sept. 23, 2010**

Data file: forams.xlsx

Software: The data file and links to the non-Office software below can be found on Sakai.

* Excel
* PAST (only available for PC):
	+ <http://folk.uio.no/ohammer/past/>

http://sepmstrata.org/thinsections/caco3-skeletal.html

* + On the right side, click on “Download”, then save the file “PAST.exe”. This is a *free* statistics program that does basic stats as well as paleo-specific stuff. Documentation, including a users guide, is also on this website.
* Analytical Rarefaction (available for PC or Mac)

I have given you a subset of data taken from Hammer & Harper (<http://folk.uio.no/ohammer/past/>). These are counts of various genera of foraminifera taken at 6 different ocean depths. Forams are protists that produce *tests* (skeletons) of different shapes & compositions – see above. Because they make hard parts, they are prevalent in the fossil record and are great indicators of water depth, temperature, and deposit age.

This assignment will walk you through how to calculate the diversity indices that we talked about in class, as well as how to compare indices.

Be sure turn in answers to all of the questions in **bold**. Assignments should be typed/computer printouts.

1. Calculate the total number of individuals within each sample (n).
2. Calculate S for all samples (depths).
3. **Calculate β for a comparison between the sample taken at 151-200 ft and 251-500 ft. Then do it for 151-200 ft and 51-100 ft. Which is more similar according to β?**
4. If you haven’t already done so, put all of your data in columns (*copy -> paste special -> transpose* is fastest), that you have genera in rows and samples (depths) in columns.
5. Open up PAST – a blank spreadsheet should appear. Copy and paste *only the data* into the PAST spreadsheet. If you want to put labels on your data (depths, etc.), go to *Edit ->Rename rows* or *Rename columns.*
6. Select the first column by clicking on the column header. Then go to *Diversity ->Diveristy indicies*. Record the Dominance\_D (this is Simpson’s D), Shannon\_H, and Equitability\_J (this is Pielou’s E) values. Calculate D, H, and E for all of the samples.
7. **Make a table with the following information for each sample: n, S, D, H, and E.** **Which sample has the highest S? Is it the same one that has the highest D? Highest H?**
8. Now we will statistically compare the indicies without worry about rarefaction. This requires a little data manipulation. PAST will only allow you to compare columns that are next to each other, so you will have to do some copying and pasting to compare all of the samples. To do so, select two adjacent columns, then go to *Diversity -> Diversity t-test*. If you’ve been reading your course packet, then you should know how a t-test works (how to figure out if something is different or not). While this t-test is a little different than a standard t-test, reading the results is exactly the same. **Do all pairwise comparisons of diversity and present the data in a table (t and p-value). Which samples are the same, and which are different?**
9. Next, we’re going to use rarefaction to adjust everything to the same sample size, then compare our results again. While you can do rarefaction in PAST, I like Steve Holland’s program, which can be found for free at: <http://www.uga.edu/~strata/software/index.html> . Scroll down to “Rarefaction” and pick the “Analytical Rarefaction” program. Note that you have to choose the correct icon for your operating system (pre-OS X, OS-X, and Windows). The following directions are adapted from Dr. Holland’s instructions – note that there is an instruction webpage that goes along with this software.
10. Each species will need its own data file, which will consist only of 1 column of integers representing the raw abundance data. Do not put any other info in there! Data files should be made in a program such as NotePad (Windows), TextEdit (Mac), or SimpleText (Mac). So, if we had 5 species in a sample whose abundances are 5, 87, 6, 3, and 12, the data file would look like this:

5

87

6

3

12

1. Save the data file and put it in the same folder as the Analytic Rarefaction program. On a Mac, the file should be \_\_\_\_\_.dat , but on a PC it needs to be \_\_\_\_\_\_.txt . Make sure that there are no spaces or funky characters in your file names.
2. Double click on the Analytic Rarefaction icon. The program will ask you for the name of the data file. Be sure to enter the name with the .txt or .dat extension (so you would type *name.txt*). You will then be asked how frequently the program should do the rarefaction calculations. Since we’re working with <200 specimens, we’ll go ahead and do 5. The program will then do the rest, including generating a file called rarefaction.res . **NOTE**: this is always the name of the file it generates, so you need to rename it before you run the program again, otherwise it’ll overwrite it. Since the data is automatically saved to the .res file, you can just close the program when you’re finished. Repeat these steps for each sample.
3. Next, open Excel. Then open one of the rarefaction files that you just made. The Text Import Wizard should pop open. Select “delimited”, then hit “Next”. Click the box next to “Space” and hit “Finish”. You now have several columns of data: n, E (this is not evenness! This is the expected number of species, or expected S), Var (the variance of E), Upper95, Lower95, Upper99, and Lower 99. The last for variables are the upper & lower 95% and 99% confidence limits.
4. Get all of your generated data into one excel workbook. It’s best to keep each sample’s data on its own tab.
5. We’re going to use the 95% confidence limits (CLs), but they need to be tweaked first. The rarefaction program gives you the absolute values of the CLs. To make them work with Excel’s error bars, we need to make the relative to the expected S. Set up two new columns – one for the upper and one for the lower. To recalculate the upper CL, subtract the expected S from the upper CL. To recalculate the lower CL, subtract the lower CL from the expected S.
6. Time to make the graph, just like you’ve seen in lecture. Put N on the x-axis and E on the y-axis. Graph each sample on the same graph. Once you have all of your samples on the same graph, draw a vertical line at the smallest n (shortest rarefaction line). **Print this graph, then make a table with the expected diversities for each sample at this sample new size. How do these values compare to the S values that you were getting from the unrarefied data?**
7. Next, we’ll add in the error bars using those new confidence limits you calculated. To do so, click on one of the lines you’ve made. Then go to Chart Tools -> Layout -> Analysis -> Error Bars -> More Error Bar Options. Make sure Display Direction is Both, then under “Error Amount”, click Custom and Specify Value. Selecting the values is the same as it is when you’re selecting data for the graph. For the Positive Error Value, select the recalculated upper CLs, and for the Negative Error Value, select the recalculated lower CLs. Excel may automatically generate horizontal error bars – you can delete those by clicking on them and hitting delete. **Add the error bars for all of your curves and print this graph.**
8. Time to compare your curves! While we can’t use formal statistical testing at this point, because we have no replicates, you can use the error bars. **Using the criteria set forth in lecture, which samples may be different? Which may not be? Are there any that you can definitely make a decision on? Does this match what you got from the unrarefied data?**
9. Last thing to do is to make rank-order abundance (RAD) plots of your original data. You should be able to do this easily in Excel. First, for each sample, sort your data from most abundant to least abundant. Assign the rank of “1” to the most abundant, “2” to the next most, etc. Then, for each genus, calculate the *relative abundance* (# individuals/n). Make a scatterplot with rank on the x-axis and proportion on the y-axis. **Print this graph. What does this graph suggest about evenness for each sample? Does it match what you found earlier? Why or why not?**