Math151 at the University of Tennessee, Knoxville - Chat for August 24, 2015 with the course instructor, Louis Gross.

I will be online starting at 7:30PM and will be happy to answer questions regarding any aspect of the course, assignments, etc. You can type in this document to ask questions - note that you need to be logged into your UTK Google Drive account to be able to type in this.

When you ask a question, please do not use your name because this document will be saved and publicly posted after we close it. I will be on-line at least until 8:30PM but will stay on longer if there are still questions.

Will we be given an outline for what exactly we need to study for each exam?

I will be giving out a sample exam prior to each exam and we will plan extra sessions, in the evening if necessary, to answer questions. Regarding coverage for each exam, the main expectations are that you will be able to do problems similar to the ones assigned from the text. Lou

Note as well that on the home page there is a set of comments on STudy Skills that I encourage you to read through and follow.

Okay great! Thank you so much!

After class today one of the attendees asked me about the meaning of “diversity” in the measures described in the text. Essentially this term (you may have heard the term “biodiversity” for this) is a formal mathematical way to calculate not just how many different species are present in a region (this is the common language meaning of the term) but also how these species are distributed. The number of different species is called “richness” and assessing how equally the species are distributed is called “evenness”. So if a region has an exactly equal number of individuals in each of the species present then it is a region with high evenness. The text gives one measure of diversity - the Simpson index
here given as

$$1-D=1-\sum_{i=1}^{S}\frac{n_i(n_i-1)}{N(N-1)}$$

A good way to think about this is to consider a couple of extreme cases

a. Suppose each species has exactly the same number of individuals present so in that case each of n_i is equal to the same value - N is the total number of individuals of each species. Then $N = Sn_i$ and since each n_i has the same value and as the number of species S gets large, D gets close to $\frac{1}{S^2}$ and so the SID gets close to 1.

b. If on the other hand, there is one species which has a very large number of individuals and all the other species in the region have a very small number of individuals, then most of the terms in the sum are near zero except for the species which is superabundant and for that species the term in the sum is near 1, so the SID is near zero.

Lou

Can you please explain dividing data sets into quartiles if the data set contains an odd number of data ie, 17?

The first step is to find the median, which divides the data into two equal groups if there are an even number of data points, and into two equal groups not counting the median if there are an odd number of points. To do this you rank the data in order from smallest to largest and choose the value of the point in the middle - in the case of 17 data points this would be the 9th data point on the list, so there are eight points larger than this (or possibly equal to it) and eight data points smaller. Then to find the quartiles, you do exactly the same procedure with each of the two parts of the data set on each side of the median. Thus in this case the first quantile will be all the points up to the arithmetic average of the two "middle points" in the first half of the data set - e.g. the first quartile goes up to the average of the 4th and 5th data points on your ranked list and then the second quartile goes from the average of the 4th and 5th data points up to the median. You then do the same process for the second half of the data to find the third and fourth quartile.

Lou

So, in the case of a 17 point data set that was sequential - 1-17 - the first quartile would be 1-4.5, the second 4.5-9, and so on? Or do I have that confused?

You have it exactly correct! Note that this becomes a bit more complicated when you are finding quintiles or other "odd" splits - there are some differences between different methods on getting exactly the splitting values. You can see this becomes real problem if there are fewer data points than the number of "tiles" you are trying to split the data into!

Lou

But, the base idea is that each split is equal, right? So, it is all a matter of making sure that each split contains the same quantity of "stuff"?

Yes, that is exactly the idea. It only becomes a problem when the number of data points is not divisible by the number of "tiles". In practice, folks generally don't use more than quartiles much,

unless there is a very large number of data points in which case we can get any %-tile we want without the difficulties of exactly how to calculate the splits.
Lou

Going back to the diversity stuff for a minute… My brain keeps wanting to qualify the result of 1 - D as a percentage. So, if the equation spits out 0.8974, my brain says, “Oh, this is 89.74% diverse!” ...which is okay in that the number represents greater diversity as it nears 1, but I have to constantly tell myself it doesn’t mean a percentage. But, does the number mean anything other than a “symbol” for the diversity?

Actually, you have it very close to correct only we haven’t discussed probabilities yet, which your “percentages” are close to. So the SID of .8974 is equivalent to saying that if you choose two individuals at random (think of a big pot that contains all the individuals of any species in the region, from which you choose two without looking) there is a .8974 probability that they are from different species.
Lou

Not any two particular species - just *different* ones?

Yes - any two different species. You can imagine that you repeat this “choosing two individuals” a very large number of times - say 10,000. Then SID = .8974 would say that about 8974 out of the 10,000 times you do this, the two individuals chosen are different species, and the remaining 1026 times the two individuals will be from the same species.
Lou

Great - thank you for explaining that.

I am signing off now - everyone have a good evening.
Lou