**Name \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Overview**: When confronted with a question about the past, paleontologists often look to museum collections for gathering the raw data necessary for their studies. But how accurate is that data? How confident are we that materials we are looking at represent the past? There are numerous filters that our collections go through. In this assignment, it is our task to unpack some of those filters.

**Activity A:** Comparison of modern and ancient ecosystems

 To start, let’s think about what we mean by comparison.

1. Write down 5 different situations in which you want to compare two items. For example, you are at the store and are considering two brands of apples. What metric do you use to select a “winner”?

|  |  |
| --- | --- |
| **Pairs of items** | **Metric for “winning”** |
| *Example: granny smiths vs. honeycrisp* | *Amount of sweetness* |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

1. Of your metrics above, which fall into the following categories? Might be more than 1.
	1. Numerical comparisons (scores, measurements, etc.)
	2. Calculated comparisons (averages, sum of several quantities, etc.)
	3. Pass/fail comparison (a threshold)
	4. Graphical comparisons (looking at visual differences on a graph)
	5. Statistical comparisons (t-test or anything with a p-value)
2. If there is a category that you did not have in your initial list, what are some examples of comparisons in that category?
3. Thinking about different comparison methods, which (at this point) “feels” the best in a scientific context? Which “feels” the worst?
4. Which do you think are used in the context of comparing diversity in ecosystems? Why?

Let’s look at some data. Below is a table of organisms from the Mediterranean Sea. These organisms were observed in artificial reef settings over a 6-year time span (N = 281) by a group of volunteer scientists. The vertebrate values came from a different study.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Algae | 23 |  | Fishes (including sharks) | 51 |
| Ascidians | 16 |  | Mollusks | 43 |
| Brachiopods | 1 |  | Phoronidians | 1 |
| Bryozoan | 17 |  | Plants | 1 |
| Cnidarians | 24 |  | Sponges | 17 |
| Crustaceans | 18 |  | Vertebrates | 43 |
| Echinoderms | 13 |  | Worms | 13 |

And here is a dataset of fossils from Malta (an island in the Mediterranean Sea) during the Miocene (~23 mya to 5.3 mya, N= 345). The dataset was collected from the Paleobiological Database.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Algae | 55 |  | Fishes (including sharks) | 66 |
| Ascidians | 0 |  | Mollusks | 54 |
| Brachiopods | 6 |  | Phoronidians | 0 |
| Bryozoan | 11 |  | Plants | 2 |
| Cnidarians | 11 |  | Sponges | 3 |
| Crustaceans | 29 |  | Vertebrates (Cetaceans, Pinnipeds, Archosaurs, and Sirenians) | 63 |
| Echinoderms | 40 |  | Worms | 5 |

1. Just looking at the raw data, can you say with certainty that these datasets are the same or different?
2. What makes comparing the data in this manner difficult?

***Diversity Metrics***

What we need is a way to summarize all the data into a single number that can be compared across the two datasets. We want to summarize the biodiversity of the two settings, and if we determine that one is more diverse than the other, that would suggest the two are different. Of course, there are multiple reasons why they could be different. Let’s rule out a few of those reasons.

1. Below are reasons why a biodiversity metric could show two populations are different. State whether you think this is a valid reason for this particular case or not.
	1. They could be different because the locations had drastically different conditions like temperature and salinity.
	2. They could be different because different people collected the data.
	3. They could be different because one has more samples in it.

Other issues could be that the modern samples were collected over a period of 6 years, while the ancient samples were collected from samples that were deposited over 18 million years of time. There are some complex statistics that can be used to remove this bias, but that is beyond the scope of this activity. We will ignore that for now.

1. What would you expect the result of this comparison to be if the fossil record was inaccurate? Write a hypothesis (use an if/then statement) about the potential results of the biodiversity metrics and accuracy of the fossil record.

***Let’s calculate a diversity metric:*** One diversity metric is called the Berger-Parker Index. It is a measure from 0 to 1 and shows the “balance” of the ecosystem. Values close to 0 would mean it is very balanced where nearly every organism occurs the same number of times (for example, 2 monkeys, 2 gorillas, 2 orangutans). 1 means it is unbalanced, which means a lot of one organism (e.g. 6 monkeys, 0 gorillas, 0 orangutans). Nmax is number of the most common organism.

Berger-Parker Index = Nmax / Ntotal

1. Calculate the Berger-Parker score for both ecosystems. Are the scores similar?
2. How satisfied are you that this is a good method of measuring diversity? Please explain.

Berger-Parker is ok as a metric. It is super easy, but has been show to be inaccurate in situations that have a larger number of rare organisms since it basically ignores them.

1. What is meant by the above statement? Specifically, the “ignores them” part?

 Another diversity metric is Shannon’s Diversity. It has similarities to Berger-Parker, but repeats the calculation for each group of organisms and adds up the scores. Large score, more diversity.

H’ = Shannon’s diversity

pi = proportion of species (N for organism / Ntotal)

ln = natural log function

∑ = symbol for “sum of”

1. Find the Shannon’s diversity of both datasets. Do this on your spreadsheet (Tab 3).
2. How satisfied are you that this is a good method of measuring diversity? Please explain.
3. Are the two populations different? Explain.
4. What makes biodiversity metrics difficult to use as a comparative method? What makes them easy to use as a comparative method?

***Hypothesis Tests***

 As you may have noticed, using diversity metrics still leaves you with the same problem as when you compared the raw numbers. But now, instead of a table of numbers, you have just one number and it is hard to know how much of a difference in numbers is enough. Is 1.1 different enough from 1.2? Maybe. What we need is a statistical test; one that gives us a single yes/no answer.

 You have probably heard of hypothesis tests before in the context of p-values. A p-value of 0.01 simply means that there is a 1% chance that two samples are the same. For most people, if the odds are under 5% we say that they are probably not the same samples.

 There are lots of hypothesis tests to use, but we are going to use one called χ2 (chi-squared). It compares the **E**xpected results vs. the **O**bservations you made. This generates a chi-squared value which corresponds to a p-value.

1. After calculating the chi-squared, do you believe these samples are different? Explain your results. (Tab 4)
2. What makes hypothesis tests difficult to use as a comparative method? What makes them easy to use as a comparative method?
3. In your own words, how is this comparative method different from the previous one?

**Activity B:** Comparison of modern and ancient ecosystems while controlling for environment

 Check out the cleaned data tab in the spreadsheet. You might have noticed in the dataset that not everything came from the same environment. In particular, **sandstone** implies a very shallow water environment, for example a beach. While **limestone** implies a bit deeper water environment, such as offshore in a reef environment. Perhaps part of the reason we got the answer we did was because we mixed a bunch of environments together for the ancient dataset.

1. Why would should we choose to limit our analysis to only those in limestone (aka reef environments) when comparing the ancient to modern samples?
2. Why would mixing environments cause the diversity to be higher or lower?
3. Make a prediction: what will happen to Berger-Parker and Shannon’s values if we do the same analysis above, but limit it to only those samples found in limestone?

 Using the same data set as before, make a new copy of the data but with only samples found in limestone or reefs. Copy rows with these terms into the new tab in the spreadsheet.

1. What is the new table of results for the ancient environment? Use COUNT function and record the numbers in the table below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Algae |  |  | Fishes (including sharks) |  |
| Ascidians |  |  | Mollusks |  |
| Brachiopods |  |  | Phoronidians |  |
| Bryozoan |  |  | Plants |  |
| Cnidarians |  |  | Sponges |  |
| Crustaceans |  |  | Vertebrates |  |
| Echinoderms |  |  | Worms |  |

1. What is the Berger-Parker and Shannon’s diversity for the new dataset?
2. Compare your new dataset to the original modern dataset. Based on a χ2 are your two samples the same or different?
3. Is the fossil record an accurate representation of biodiversity? What caveats are there?
4. Which organisms have a better fossil record? Why might this be?
5. After the first two activities, what can you say about natural filters in the fossil record? That is, what are some challenges of directly comparing the fossil record to modern environments?

**Activity C:** Comparison of museum collections to the fossil record

 There is another level to this study, and that is museum collections. Museums are run by humans, and we all have preferences for what we study. For example, the largest scientific organization for paleologists is the *Paleontological Society*. It is made of largely invertebrate paleontologists and has around 1,500 members. Another organization, *Society of Vertebrate Paleontology*, has around 2,300 members. Even though vertebrates make up less that 1% of known organisms, the membership number for the *Society of Vertebrate Paleontology* is higher. This forces us to ask the question, are museums an accurate reflection of the fossil record, or are they biases because of our preference for dinosaurs, mammals, etc.

1. First, make a prediction. What kinds of organisms do you think are most common in museum fossil collections. Why?
2. What are negative repercussions of your answer above?

Go to [American Museum of Natural History](http://research.amnh.org/paleontology/search.php) (AMNH) and the [Natural History Museum of London](https://data.nhm.ac.uk/search) (NHM) online collection repositories. In their respective search bars, search for the number of records (i.e., specimens at those institutions) for the following, plus 5 of your choosing. Do the same with the [Paleobiological Database](https://paleobiodb.org/navigator/) (PBDB). In each, limit your search to specimens from the Cenozoic. Instructions are included on spreadsheet in the Activity C tab.

|  |  |  |  |
| --- | --- | --- | --- |
| **Name of Organism** | **AMNH** | **NHM** | **PBDB** |
| Cnidaria |  |  |  |
| Sponges (Porifera)  |  |  |  |
| Bryozoa |  |  |  |
| Horse (Equus) |  |  |  |
| Turtles (Testudines) |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

1. Investigate a few of the records in the two museum collections. What kind of information is provided?
2. Run a chi-squared test for your museum collections with PBDB as the expected results (you will need to two comparisons). What are the results?
3. Are the datasets different or the same? Why do you think that is?
4. Do you think museum collections are like the filters talked about at the end of activity B? Is one more or less concerning to you when trying to study the fossil record?