Multivariate Statistics (**MARS6300**) - Homework 5 **Name:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Distributed:** Tues, March 13, 2018 **Due:** Weds, March 22, 2018

**Instructions:** Copy and paste your answers below and turn in a word file and two excel files by the end of due day via email to khyrenba@gmail.com. Please use email title “MARS 6300 hw#5” and label all files with you’re a suffix including your name (e.g., MARS6300\_hw5\_hyrenbach). Unlabeled emails / files will be penalized 10% of points.

**The objectives of this homework are:**

A) To critically evaluate the “biological indicator” literature.

B) To perform and interpret Weighted Averaging Analysis.

C) To perform and interpret Polar Ordination Analysis.

To complete this homework, you will need:

* Instruction file: “MARS6300\_hw5.doc” (open with word file) – **Turn this file in**
* “Biogeography\_data.xls” data file: (open with excel) – **Turn this file in**
* Two pdfs: Dufrene & Legendre 1997, McGeoch & Chow 1998
* “Biogeography\_environment.xls” & “Biogeography\_samples.xls”
* “Biogeography\_weights.xls” is a template, to show you how to organize the file.

(Feel free to modify this file with your data and use in PC-ORD).

1. **Reviewing Indicator Species:**

Read Dufrene & Legendre 1997, and McGeoch & Chow 1998 and answer the following questions:

* Define “specificity” in plain words
* Provide an equation for calculating “specificity”
* Define “fidelity” in plain words
* Provide an equation for calculating “fidelity”

Discuss how these three ways of integrating the “fidelity” and the “specificity” metrics will affect the resulting indicator value: (i) sum (specificity + fidelity) , (ii) multiplication (fidelity \* specificity), (iii) sqrt (fidelity \* specificity), and (iv) ratio (specificity / fidelity). To inform your discussion, consider these two series of specificity and fidelity values, listed below:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **fidelity** | **specificity** | **SUM** | **PRODUCT** | **SQRT (PRODUCT)** | **RATIO** |
| **0** | **0.25** |  |  |  |  |
| **0.25** | **0.25** |  |  |  |  |
| **0.5** | **0.25** |  |  |  |  |
| **0.75** | **0.25** |  |  |  |  |
| **1** | **0.25** |  |  |  |  |
| **0** | **0.5** |  |  |  |  |
| **0.25** | **0.5** |  |  |  |  |
| **0.5** | **0.5** |  |  |  |  |
| **0.75** | **0.5** |  |  |  |  |
| **1** | **0.5** |  |  |  |  |
| **0** | **0.75** |  |  |  |  |
| **0.25** | **0.75** |  |  |  |  |
| **0.5** | **0.75** |  |  |  |  |
| **0.75** | **0.75** |  |  |  |  |
| **1** | **0.75** |  |  |  |  |
| **0** | **1** |  |  |  |  |
| **0.25** | **1** |  |  |  |  |
| **0.5** | **1** |  |  |  |  |
| **0.75** | **1** |  |  |  |  |
| **1** | **1** |  |  |  |  |

* What is the problem with the SUM algorithm?
* What is the problem with the PRODUCT algorithm?
* What is the problem with the SQRT(PRODUCT) algorithm?
* What is the problem with the RATIO algorithm?
1. Next, you will apply these concepts to three species lists from distinct biogeographic domains: “subtropical”, “convergence” and “sub-Antarctic”.

For each biogeographic domain, identify those species that would be valuable indicators using both the “specificity” (density) and the “fidelity” (presence / absence) concepts. Hint: To do this, combine the information from all three domains, so you consider the abundance and the occurrence of each species across all three domains at once.

Explain below how you would do this:

Use this approach to calculate the indicator values for each species in each domain. Report the “indicator values” for each species below:

Use the column / row summary tool in PC\_ORD and calculate column / row

totals for the data in matrix 1. Paste the three sets of excel file cells below, showing the species list, specificity value, fidelity value, and overall indicator value for all three biogeographic domains:

* **Subtropical:**
* **Convergence Zone:**
* **Sub-Antarctic:**
1. **Perform weighted averaging analyses:**

Using the overall indicator values you calculated, create three matrix2’s: one for each biogeographic domain. These matrices will show the weights for each species as indicators of each biogeographic domain.

Load the sample data “biogeography\_samples.xls” and perform three Weighted Averaging Analyses, using the three matrix 2 “weights” you just created.

For each analysis, paste a plot of the samples arranged along a 1-D axis (Note: ask for 2-D plot), showing the abundance of the best indicators for that biogeographic domains (based on your answers from section 1).

**Final question:** based on the results of the three weighted averaging analyses, classify each sample as belonging to one of the three biogeographic domains, below. For each “bin” sample, show its score in the three analyses you run and then use this information to classify the sample as one of the three biogeographic domains:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Subtropical | Convergence | Sub-Antarctic | **Classification** |
| bin1 |   |   |   |  |
| bin2 |   |   |   |  |
| bin3 |   |   |   |  |
| bin4 |   |   |   |  |
| bin5 |   |   |   |  |
| bin6 |   |   |   |   |
| bin7 |   |   |   |   |
| bin8 |   |   |   |   |
| bin9 |   |   |   |   |
| bin10 |   |   |   |   |
| bin11 |   |   |   |   |
| bin12 |   |   |   |   |
| bin13 |   |   |   |   |
| bin14 |   |   |   |   |
| bin15 |   |   |   |   |
| bin16 |   |   |   |   |

1. **Perform Polar-Ordination analyses:** Open the “biogeography\_environment.xls” file and use this information to test three hypotheses about the organization of the samples: (H1): water temperature (SST) is most important ecotone, (H2) productivity (CHL) is most important ecotone, and (3) latitude is most important ecotone.

To make this test, perform three Polar Ordination (Bray Curtis) tests, in the Ordination menu, using these default settings, and asking for 1 subjective axis:



1. SST: Identify the two poles for this analysis:

“warm bin”: “cold bin”:

Perform analysis and paste the following results, below:

* Report amount of variance for axis 1 (objective):
* Report amount of variance for axis 2 (subjective):
* Report amount of variance for axis 3 (subjective):
1. CHL: Identify the two poles for this analysis:

“highest bin”: “lowest bin”:

Perform analysis and paste the following results, below:

* Report amount of variance for axis 1 (objective):
* Report amount of variance for axis 2 (subjective):
* Report amount of variance for axis 3 (subjective):
1. Latitude: Identify the two poles for this analysis:

“northern-most bin”: “southern-most bin”:

Perform analysis and paste the following results, below:

* Report amount of variance for axis 1 (objective):
* Report amount of variance for axis 2 (subjective):
* Report amount of variance for axis 3 (subjective):

 **Final question:** Which analyses gave the best result?

* Which of the three “objective” axes performed better – higher variance explained?
* Which analyses explained the most variance overall?
1. **Synthesis:** To finish, calculate the weighted average scores for the pole “samples” you used in the three Polar Ordinations. Using the proportional species composition of the samples and the indicator values you calculated in question 1, calculate the following:
* **SST hypothesis:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Subtropical** | **Convergence** | **Sub-Antarctic** |
| **“warm” pole bin** |  |  |  |
| **“cold” pole bin** |  |  |  |

* **CHL hypothesis:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Subtropical** | **Convergence** | **Sub-Antarctic** |
| **“highest CHL ” pole bin** |  |  |  |
| **“lowest CHL” pole bin** |  |  |  |

* **LAT hypothesis:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Subtropical** | **Convergence** | **Sub-Antarctic** |
| **“northern-most” pole bin** |  |  |  |
| **“southern-most” pole bin** |  |  |  |

Discuss (qualitatively) whether the poles you identified agree with the species indicator values you calculated? Hint: do subtropical species indicate warm / low CHL / low LAT poles? What environmental poles do the sub-Antarctic species indicate?