Data Analysis & Statistics for Earth Scientists
Nanjing University, Spring 2026

Lab 4 Assignment

1. The Conodont dataset (Conodonts.csv, Conodonts.dat) contains counts of 10 different Carboniferous conodont taxa collected from a sequence of stratigraphic units in western Kansas. These units represent component parts of four megacyclothems into six characteristic depositional facies. It has been suggested that conodont species partitioned the water column into depth zones, much like modern nektonic species do today. If so, conodont taxa assemblages should be found in particular depth-related lithofacies. This hypothesis can be tested using singular value decomposition (SVD) to analyse both the *R*-mode and *Q*-mode structure of these data. In ecological data analysis SVD is often referred to as correspondence analysis. Compare and contrast the *R*-mode and *Q*-mode structure of the conodont dataset to complete the following assignment. (150 points)

    a. Do these data require any transformation prior to either analysis? Explain your reasoning. (10 points)

    b. Perform correspondence analysis on these data (or some transformed aspect thereof).

        1. Decide how many components to retain. Explain your reasoning. (20 points)

        2. Construct a biplot (or set of biplots) of the conodont species ordination and the lithofacies ordination. (20 points)

        3. Interpret the biplot axes in terms of the alignment with particular conodont species and particular lithofacies. (20 points)

        4. Is any clustering – either of conodont species or lithofacies – evident in this/these ordination(s). (20 points)

        5. Which conodont species are most closely associated with which lithofacies. (20 points)

        6. Given the inferred depths of the megacyclothem lithofacies attempt the reconstruction of a depth zonation for the conodont taxa. (40 points).

2. The Carnivores (Distances) dataset (Carnivores (Distances).csv, Carnivores (Distances).dat) contains linear distances measured on a variety of carnivore cranial features. You may assume all distances were measured in the same units. Compare and contrast the information extracted by principal components analysis (PCA) and canonical variates analysis (CVA) in the manner you address the following assignments. (200 points)

    a. Do these data require any transformation prior to either analysis? Explain your reasoning. (10 points)

    b. Perform a PCA on these data (or some transformed aspect thereof).

        1. Which basis matrix is appropriate to use in the analysis of these data? Explain your reasoning. (10 points)

        2. Decide how many components to retain. Explain your reasoning. (10 points)

        3. Provide an interpretation of the PC components you retained in terms of the original variables. (20 points)

        4. Show the ordination plots for the components you retained. (10 points)

            a. Identify any clusters of taxa evident in your ordination space(s). (20 points)

            b. Use the interpretation of your space(s) to characterize any within-cluster similarities (20 points) and any between-cluster differences. (20 points)

c. Perform a CVA on these data (or some transformed aspect thereof) based on the suborders Canifomia, Feliformia, Matetheria. Note: you may use either the raw data or the PCA scores from the previous analysis.

1. Provide an interpretation of the canonical variates in terms of the original variables. (20 points)

2. Show the variate ordination plots. (10 points)

   a. Identify any clusters of taxa evident in your ordination plot(s). (20 points)

   b. Use the interpretation of your space(s) to characterize any within-cluster similarities (20 points) and any between-cluster differences. (20 points)

d. Compare your PCA and CVA results

1. Are the same clusters evident in both analyses? (20 points)

2. What differences (if any) are there in the character of the ordination spaces produced by the two analyses? (20 points)

3. What types of scientific questions can each of these analyses provide answers to? (40 points)

3. The Carnivores (Landmarks) dataset (Carnivores (Landmarks).csv, Carnivores (Landmarks).dat), consists of landmarks collected from Carnivore crania in dorsal view – the same specimens the linear distances used in the previous assignment were collected from. We'll use these to compare and contrast the traditional approach to morphometrics (assignment 2) with the newer geometric morphometric (GM) approach. Using PAST, or some other software (plenty exist) align and analyze these data according to the directions below. (160 points)

   a. Plot all coordinate datasets. (10 points)

   b. Align the landmark configurations using the Procrustes procedure. Plot an overlay of the aligned coordinates. (20 points)

   c. Calculate and create a plot of the mean coordinate configuration. (10 points)

   d. Perform a Procrustes PCA (relative warps analysis) of the Procrustes-aligned landmark configurations.

      1. Decide how many components to retain. Explain your reasoning. (10 points)

      2. Show the ordination plots for the components you retained. (10 points)

         a. Identify any clusters of taxa evident in your ordination space(s). (20 points)

      3. Calculate thin plate splines for the retained Procrustes principal components (20 points)

         a. Use these to establish a geometric interpretation of your ordination space(s). (20 points)

         b. Use the interpretation of your space(s) to characterize any within-cluster similarities (20 points) and any between-cluster differences. (20 points)

4. The Spindles dataset contains family richness data for 20 taxonomic groups sampled over 37 Phanerozoic time intervals. Use plotting routines and factor analysis to produce a three-factor analysis of diversity trends in there data, using your results to address the following issues. (100 points)

   a. Plot all 20 profiles. (10 points)

   b. Analyse the spindle shapes using Q-mode factor analysis.

1. Factor analysis differs from PCA and PCoord in that it usually requires the user to specify how many influencing factors they expect to the structuring the data. For comparative purposes (see below), we'll assume this is three. What proportion of the total distance variance is being represented by the three factor axes? (10 points)

    a. If it were up to you would you have selected a different factor number? Explain your reasoning. (20 points)

2. Provide an interpretation of the retained factors showing the data-analysis results that support your interpretation. (10 points)

3. Provide an interpretation of the richness history profiles represented by each of the retained factors noting any subgroup-level structure that exists in these data. (20 points)

c. In this exercise you are, in essence, repeating the analysis first published by Sepkoski (1981). Compare and contrast your results with those obtained by Sepkoski noting areas of agreement and disagreement. (30 points)

5. The NW Kansas dataset contains well log and core data collected from 15 levels in a well drilled in northwestern Kansas, including gamma-ray intensity measurements, sonic transit times, electrical resistivity, permeability, porosity, oil saturation and water saturation. Compare and contrast the results of a principle coordinate analysis and non-metric multidimensional scaling analysis of these levels. (80 points)

    a. Which similarity-dissimilarity index is appropriate to use in the PCoord analysis of these data? Explain your reasoning. (10 points)

    b. Do these data require any transformation prior to PCoord analysis? Explain your reasoning. (10 points)

    c. Calculate and plot the PCoord ordination using the first two PCoord eigenvectors. (10 points)

    d. Calculate and plot the two-axis NL-MDS solution for these data along with the summary stress residuals. (20 points)

    e. Compare your PCoord and NL-MDS results in terms of (i.) comparative ordination configurations, (ii.) information loss as a result of dimensionality reduction and (iii.) interpretation of the two sets of ordination axes. (30 points)