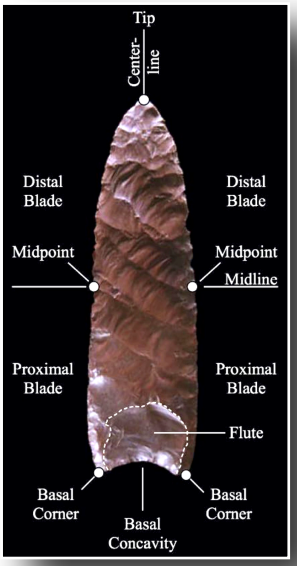
Data Analysis & Statistics for Earth Scientists
Nanjing University, Spring 2026

Lab 5 Assignment

1. The Clovis dataset (Clovis.csv, Clovis.dat) contained set of outline semilandmarks for 148 Clovis archeological points collected from four regions in the continental United States. In order to understand human cultural development archaeologists often wish to quantify and compare the forms of artifacts fashioned by craftsmen in different geographic regions, especially in those cases where written records are not available. Analysis of the nominal artifacts that define fluted projectile points, such as those characteristic of the North American Clovis culture (ca. 11,600 –10,800 radiocarbon yrs BP, 13,550 –12,850 calibrated yrs B.P.), represents an interesting case in point. Identified primarily by their production of a distinctively shaped bladed tool, the PaleoIndian cultures that created these artifacts spread across North America in the interval between the retreat of the northern glaciers at the end of the Pleistocene and the start of the Younger Dryas cold interval. Regional variation in the forms and/or shapes of these artifacts (Fig. 2) has been cited as evidence that Clovis people did, and did not vary their design in response to local environmental conditions or situations, including fabrication material availability, prey types, cultural differences, use in ceremonial activities, etc. Analyse the shapes of these Clovis point outlines and address the following questions. (100 points)



   a. Do these data need to be transformed prior to analysis? (10 points)

   b. Can the dimensionality of these data be reduced without losing potentially important information content? If so, by what means and by how much? Describe the methods and how the data-analysis results that support your (written) interpretation. (20 points)

   c. What are the major directions of shape variation across the pooled sample? Describe the methods and how the data-analysis results that support your (written) interpretation. (20 points)

   d. Do distinctions, in terms of Clovis-point shape variation, exist between the regions from which the samples were collected? Describe the methods and how the data-analysis results that support your (written) interpretation. (30 points)

   e. If any regional distinctions exist, are they significant statistically? Describe the methods and how the data-analysis results that support your (written) interpretation. (20 points)

2. Calibrating airborne and satellite remote sensing systems requires extensive “ground truthing” via comparison with features whose local and/or regional distributions are known. As part of a calibration exercise a high-angle arial photograph was taken of a test area in southern Arizona in which the taxonomic identification of each individual plant was known. These plants were classified into two types based on their ecologies: extremely drought-tolerant creosote-type plants (e.g., *Larrea tridentata*, left)) and moderately drought-tolerant brittlebush-type plants (e.g., *Encelia fairnose*, right). The locations of these plants in the test area are given in the Arizona datasets (Arizona.csv, Arizona.dat). Use quadrate analysis to quantify and compare the distributions of these plant types in answering the following questions. (90 points)

   a. Do these data need to be transformed prior to analysis? (10 points)

   b. Use the combined distributions of both plants to define 9 cell, 16 cell and 25 cell quadrate sampling schemes. Show your cell counts. (10 points)

c. Test for distributional uniformity for the drought-tolerant creosote-type plants and the brittlebush plants using each sampling scheme. Show the data-analysis results that support your (written) interpretation. (20 points)

d. Combine the two datasets and test for distributional randomness and clustering for the drought-tolerant creosote-type plants and the brittlebush plants using a 15 x 15 quadrat sampling scheme. Show the distribution map and data-analysis results that support your (written) interpretation. (20 points)

e. On the basis of the results you obtained in 3c and 3d decide whether the two plant types exhibit similar of different distributions in the test area. Refer to specific data-analysis results that support your (written) interpretation. (30 points)

3. At present, over some 190 have been conformed to exist on Earth (see Earth Impact Database). The North American Impact Structures datasets (North American Impact Structures.csv, North American Impact Structures.dat) lists the locations of the 60 impact structures known from North America (the US, Canada and Mexico) along with the 10 suspected structures that remain unconfirmed at this time. Analyze the geographic distribution of these impact known and suspected structures and answer the following questions. (120 points)

a. Do these data need to be transformed prior to analysis? (10 points)

b. Are the confirmed impact structures distributed uniformly across North America? Describe the methods and how the data-analysis results that support your (written) interpretation. (20 points)

c. Do the confirmed impact structures exhibit unusual clustering or dispersion across North America? If so, what might this arise from? Describe the methods and how the data-analysis results that support your (written) interpretation. (30 points)

d. Are the unconfirmed impact structure positions distributed uniformly within their own spatial field? Describe the methods and show the data-analysis results that support your (written) interpretation. (20 points)

e. If positions of the unconfirmed structures are combined with the positions of the confirmed impact structures, would this change your answers to 3b and/or 3c? Show the data-analysis results that support your (written) interpretation. (40 points)

4. The cuticula of many trilobites display numerous tubercles , which are small raised bumps or granular-like protuberances on a structure's surface. In some trilobites these are thought to house specialized organs that allowed the trilobite to sense its local environment. If this hypotheses is correct it would be interesting to know whether the distribution of these tubercles was random, clustered or dispersed. The Tubercles dataset (Tubercles.csv, Tubercles.dat) contains information on the positions of 136 tubercles collected from the left-hand side of the cranidium from a single *Paradoxides forchhammeri* specimen. Use nearest-neighbor analysis to answer the following questions. (100 points)

a. Do these data need to be transformed prior to analysis? (10 points)

b. What is the expected mean nearest neighbor distance for these data based on the naîve estimate and Donnelly's (1978) correction for edge effects? In this calculation use the area and perimeter of the distribution's convex hull as the area estimate. (20 points)

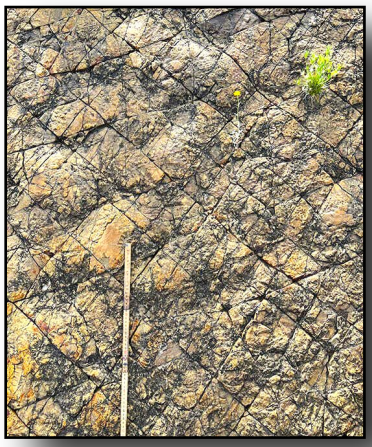
c. What is the mean nearest neighbor distance for these data? (20 points)

d. Based on the results you obtained in 4b and 4c, decide whether the tubercles in this specimen display a random, clustered or dispersed distribution. Show the data-analysis results that support your

(written) interpretation. (30 points)

e. Based on the results you obtained in 4d offer explanations for the distribution, and the function, of tubercles under the assumption that they housed some sort of sensory organ. (20 points)

5. In the development of petroleum reservoirs it is important to understand the direction(s) petroleum resources can be expected to flow. In this context it would be useful to know whether lineaments exposed on the surfaces of reservoir-rock outcrops are indicative of fractures at depth. The three Odessa datasets (Odessa North, Odessa Northwest, Odessa West) contain surface lineament direction measurements for two reservoir rock formations In the Odessa North field independent assessments of the reservoir fracture pattern characterizing the Grayburg Dolomite at depth suggest its fractures exhibit a trend of approx. N 75° E. In the Odessa Northwest field (also the Grayburg Dolomite) at-depth studies suggest fracture trends of N 35° E and N 55° W. The Odessa West field drawn petroleum from the San Andrews Limestone where water flood breakthrough data suggest a channeling trend of N 75° E. The Odessa North, Odessa Northwest and Odessa West datasets each contain surface lineament measurements for outcrops of the reservoir rocks in question. Use these data to perform the following analyses and answer their associated questions. (120 points)



a. Do these data need to be transformed prior to analysis? (10 points)

b. Display direction rose charts for the Odessa North, Odessa Northwest and Odessa West datasets. Use these to estimate the mean lineament directions and their associated 95 percent confidence angles. (30 points)

c. On a pairwise basis, do these lineament data all exhibit the same trend directions? Show the data-analysis results that support your (written) interpretation. (30 points)

d. Are each of the lineament datasets consistent with the estimates of at-depth fracture trends for their respective petroleum production fields? Show the data-analysis results that support your (written) interpretation. (30 points)

e. How might any discrepancies between surface lineament trends and at-depth direction flow trends be accounted for? (20 points)