

## Homework 4 – Due 11:59 pm CST, 13 March 2021

The total points on this homework is 80.

1. The Olive Oils dataset available on Canvas consists of measurements of eight chemical components (columns 2-8) on 572 samples of olive oil. The samples come from three different regions of Italy. The regions are further partitioned into nine areas: areas A1–A4 belong to region R1, areas A5 and A6 belong to region R2, and areas A7–A9 belong to region R3. The first column of the file provides the indicator for the nine regions. (Note that the file has a header.) Answer the following questions:
  - (a) Provide appropriate graphical summaries for the dataset. You may plot all nine regions (using color) in one figure, and also the individual sub-regions separately to better understand the dataset. Comment on the distinctiveness between the chemical composition of the olive oils in the three main regions, as well as individually between the sub-regions. [1 page; 20 points]
  - (b) We will now focus on Region R2. Answer the following questions.
    - i. Calculate the correlation matrix for the chemical composition in each of the two sub-regions. Display the correlation matrices for the two sub-regions side-by-side, and comment on possible differences. [1 page; 5 points]
    - ii. Also, compare the marginal standard deviations of the composition of the eight chemicals for the two regions. Once again, you may use parallel coordinate plots if it helps to easily compare the standard deviations for the two groups. [10 points]
    - iii. Test for differences in dispersions among the two groups. [10 points]
    - iv. For each of the two groups, is it reasonable to assume multivariate normality? [10 points]
    - v. Regardless of how the tests in 1(b)iii and 1(b)iv above turn out, test for equality of the mean chemical compositions between the two groups. Specifically, report the Hotellings'  $T^2$ -statistics and its  $p$ -value. [10 points]
    - vi. Provide also individual pairwise  $t$ -tests for the differences in the composition of the eight chemicals among the two groups, using Bonferroni and also after controlling false discovery rates, at the 5% level of significance. [5 points]
    - vii. Restricting attention to the coordinates for the fifth and sixth chemicals (and on the same plot), draw the 95% confidence ellipses for the two groups. (You may use the `ellipse` package.) [5 points]
  - (c) We will now focus on differences among the three main regions (R1, R2 and R3).
    - i. Display the three regional means of the chemical compositions for the olive oils in the three regions, using Chernoff faces. Comment. [1 page; 5 points]
    - ii. Perform a one-way multivariate analysis of variance to test for differences in mean chemical composition among the three regions. Evaluate the assumptions made in making this inference? [0 points]