Politecnico di Milano
Scuola di Ingegneria Industriale e dell'Informazione
Applied Statistics
September 6th, 2021

Problem n.1

Pinna Nobilis is the largest species of bivalves in the Mediterranean Sea. Their size is an index of how clean and unpolluted the waters are. At the Miramare protected area in Trieste, a sample of 82 specimens of Pinna Nobilis have been measured: both the height and the maximum width (both measure in centimetres) are reported in the file `pinnanobilis.txt`.

a) Identify possible clusters within the data using a hierarchical clustering algorithm (Euclidean distance, complete linkage). Provide the plot of the dendrogram and qualitatively identify the optimal number of clusters.

b) Assuming that the clusters identified at point a) have the same covariance structure, formulate a MANOVA model for the geometrical features (height and width) of the specimens of Pinna Nobilis as a function of the clustering membership. Report the formulation of the model, the estimates of the paramenters and verify the assumptions. Is there statistical evidence to state that the membership to a cluster has an effect on the mean features of the specimens of Pinna Nobilis?

c) Provide confidence intervals for the differences between the mean features of specimens of Pinna Nobilis belonging to the identified clusters. Use a Bonferroni correction to ensure a 90% global level. Use the computed intervals to comment about the differences among the clusters.

Upload your results here:
`https://forms.office.com/Pages/ResponsePage.aspx?id=K3EXCvNtXUKAjjCd8ope612LHtvIHvFEsEi2L6mhPglUQUExGMVZFVTZGRTlHRUhOSVhHR1lSUFVJSC4u`