Politecnico di Milano
Scuola di Ingegneria Industriale e dell'Informazione
Applied Statistics
January 17th, 2024

Problem 2: Asteroid shape analysis

The dataset `asteroids.txt` contains information on 150 different asteroids (assumed to be ellipsoidal), each characterized by eight quantitative features obtained through advanced astronomical observations.

These features include surface area, volume, major axis length, minor axis length, eccentricity, convex volume (volume of the 3D convex hull), equivalent diameter (diameter of the sphere of the same volume), and roundness (calculated with the formula $6\sqrt{\pi}\cdot$ area/perimeter$^{\frac{3}{2}}$, equal to 1 for a perfect sphere). The dataset also includes details about the type of asteroid.

a) Conduct a Principal Component Analysis (PCA) of the dataset, focusing on the quantitative variables only. Determine whether it is more appropriate to use the original variables or the standardized ones and proceed accordingly.

b) Report a plot of the loadings of the first two principal components and provide an interpretation. Report the scatter plot of the data along the first two principal components. Considering the categorical variable `Type`, interpret the results.

c) For asteroids of the `metallic` variety, construct a 95% confidence region for the mean of the vector whose components are the first two principal components. Describe this region by providing its center coordinates, axis directions, and semi-axes lengths. Provide a plot of the region. Introduce and assess the hypothesis of normality, which will here be tested at the 1% significance level.

Upload your results here: `https://forms.office.com/e/AYckL9ZbVj`