Politecnico di Milano
Scuola di Ingegneria Industriale e dell'Informazione
Applied Statistics
June 16th, 2022

Problem n.4

The file `listening.txt` contains daily measurements of music listening of 129 different songs along 2021. Consider a functional data analysis approach where, for each song, the measurements provided are considered as discrete sampling of underlying smooth functions.

a) Perform a smoothing of each datum through cubic *smoothing* splines (order = 3), with a smoothing parameter $\lambda = 10^2$. Specify the choice for the nodes of the splines. Provide a plot of the smoothed data and report the first 3 coefficients obtained for the first song.

b) Perform a functional principal component analysis of the smoothed data obtained at point (a). Report the variance explained along the first 5 functional principal components and the screeplot.

c) Propose a possible dimensionality reduction for the data and justify your choice. Plot the retained principal components.

d) Plot the retained principal components as perturbation of the mean, and interpret them.

e) Provide a plot of the scores along the first two principal components and comment the results.

Upload your results here:
`https://forms.office.com/Pages/ResponsePage.aspx?id=K3EXCvNtXUKAjjCd8ope6-9AS0GWf2lHjvGX24HiqFVUMlUwOUpJUkdCSjc2TzRSSjVSMENLMVZLMS4u`