Politecnico di Milano
Scuola di Ingegneria Industriale e dell'Informazione
Applied Statistics
July 7th, 2023

Problem 1: Products assortment of electronics and appliance shops

The file `products.txt` contains information about the product assortment of 100 high-tech electronics and appliance shops. Each column represents a specific product and indicates the degree of importance of the shop in supplying that particular product, expressed in an appropriate unit of measurement (higher values correspond to greater importance).

a) After having decided whether it is appropriate to use the original variables or the standardized ones, Perform the Principal Component Analysis of the dataset. Is there a clear number of principal components that must be considered? How many principal components are needed to explain at least 80% of the total variability of the dataset?

b) Report a plot of the loadings of the first 3 principal components. Provide an interpretation for the first principal component.

c) Report the biplot of the data along the first two principal components. What characterizes the shop labeled as `134`[1], according to the biplot?

d) Consider the dimensionality reduction suggested by point a).
Project a new shop with characteristics reported in Table 1 on the identified reduced space, providing its coordinates on the reference system of the principal components.

`cellphone`	1.41
`tablet`	1.14
`smartwatch`	1.02
`computer`	1.21
`drone`	1.11
`dockstation`	1.14
`fridge`	0.99
`oven`	0.73
`dishwasher`	0.94
`television`	1.04
`blender`	1.02
`hairdryer`	1.11

Table 1: Information of the new shop

Upload your results here:
`https://forms.office.com/e/G3vNbKpGh3`

[1] Which is the row index