Homework1

2024-10-07

Principal Component Analysis (PCA) and Multidimensional Scaling (MDS)

Import the data set "euroleague_23_24.csv" the player statistics of four teams taken part in Final Four of Euro League 2023-2024.

The variables in the data set are defined as follows:

No - Row Number

TEAM -Team of the Player

PLAYER - Player name

POSITION - Position of the player

GP - Games played

GS - Games started

Min - Minutes played

PTS - Points scored

X2P. - Percentage of Two-points

X3P. - Percentage of Three-points

FT. - Percentage of Free-throws

OR - Offensive rebounds

DR - Defensive rebounds

TR - Total rebounds

AST - Assists

STL - Steals

TO - Turnovers

BLK - Blocks

BLKA - Blocks against

FC - Personal fouls committed

FD - Personal fouls drawn

PIR - Performance Index Rating

1. First do the exploratory data analysis.

a) Discard the variable "No" from the data set. (1p)

b) Split variable "Min" using strsplit() function. Give the name "aux" to the output. The first element of each row will show the minutes that the player played in total. (1p)

c) Add a numerical variable to the data set named "Min 2" which shows on average how many minutes each player played in the game. (2p)

d) Check the structure of the data and assign correct type to each variable considering whether it is a categorical or numerical variable. (2p)

2. Application of PCA.

a) Apply PCA on all the scaled numerical variables in the data set by using PCA() function in FactoMineR package. Treat the categorical variables and the variable "PIR" as suplemantary variables using arguments quali.sup and quanti.sup correctly. (3p)

b) How many components should be extracted? Decide on the number of components considering eigenvalues. (3p)

c) Interpret the loadings/correlations of variables at each dimension (3p).

d) Use plot.PCA() function to show correlations between variables and the extracted dimensions. (For the variables you should use the argument choix = "var"). Plot all the extracted dimensions changing argument "axes".(3p)

e) Interpret variable plots. How can each dimension be named? (5p)

f) Show individual pilots for the extracted dimensions changing argumennt choix="ind" in plot.PCA() function. (2p)

g) Interpret the individual plots. (3p)

3. Application of MDS.

a) Apply metric MDS using Euclidean distance on scaled numerical variables. (2p)

b) Plot the data using the points on the first two coordinates using players names as label. (2p)

c) Interpret the plot.(3p)

d) Calculate gower distance including variable "POSITION" to the data matrix. (3p)

e) Apply metric MDS on gower distance matrix. (2p)

f) Plot individual plots on the first two coordinates (2p).

g) Use different categorical and numerical variables as labels so as to explain clusters that are constructed. (5p)

h) Which MDS do you think better group the individuals? Why? (3p)