SMDE FIRST ASSIGNMENT (40% OF THE FINAL MARK)

FOURTH QUESTION: PRINCIPAL COMPONENT ANALYSIS (25% OF THE FIRST ASSIGNMENT).

Import the data "data_Eurobasket_2025.xlsx" the player statistics of four teams taken part in Final Four of FIBA EuroBasket 2025.

The variables in the data set are defined as follows:

GP: Games Played

Min: Minutes played per match

FG: Field Goals

2PT FG: 2pt Field Goal

3PT FG: 3pt Field Goal

FT: Free Throws

OREB: Offensive Rebounds

DREB: Defensive Rebounds

REB: Total Rebounds

AST: Assists

PF: Personal Fouls

TO: Turnovers

STL: Steals

BLK: Blocks

"+/-": Point differential of the score while player on the court

EFF: Efficiency

PTS: Points

We aim to reduce dimension, obtain new independent variables to predict efficiency of players (EFF) by using linear regression.

a) Import the data set to R correctly and assign player names as rownames of the data frame (0.5p).

b) Check the structure of the data and assign correct type to each variable considering whether it is a categorical or numerical variable. (0.5p)

c) Create data frame only using variables "Position", "MIN", "FG","2PT.FG","3PT.FG", "FT","OREB","DREB", "REB", "AST", "PF", "TO","STL","BLK","EFF" and "PTS". Check the correlations between numerical variables and summarize your findings. (1p)

d) Apply PCA on the created data frame by using PCA () function in FactoMineR package. Use supplementary quantitative variable(s) in case required. (1p)

e) How many components should be extracted? Decide on the number of components to be extracted. (1p)

f) Interpret variable plots. How can each dimension be named? Plot all the extracted dimensions changing argument "axes". (You can use plot.PCA() function to show correlations between variables and the extracted dimensions. For the variables you should use the argument choix = "var"). (2p)

g) Interpret the individual plots obtained from extracted dimensions. Enrich your interpretation using supplementary categorical variable(s). (2p)

h) Find the best linear model to predict efficiency of players (EFF) on the principal components. Do not forget to test the assumptions and check the validity of the model. (2p)