

Name:

Neptune code:

Signature:

1. Download the file `www.inf.u-szeged.hu/~csendes/DentistData4.xlsx` open it in excel, and import its content into SPSS. Name the variables as Name, BadTeeths, Treatments, and Cost. Make sure that the measurement type of Name will be nominal, the others are scale. Select the proper missing data codes, and write them here! 2 points
2. Determine the clusters of the cases by K-means clustering ($K = 4$), and explain the obtained result! 2 points
3. Calculate a new variable Y with the Compute command according to the expression $2 * \text{Treatments} + 3 * \text{BadTeeths}^2 + 0.1 * \text{Rv.NORMAL}(0,1)$. Run the nonlinear regression to identify the model parameters ($a * \text{Treatments} + b * \text{BadTeeths}^2$)! How to understand the result? 2 points
4. Apply multidimensional scaling to determine which cases are close together, the variable Name is to be used as label! Copy here the obtained 2 dimensional plot. How to understand the result? 2 points
5. Run the Factor analysis to our data! Explain the obtained result! 2 points

Optional + problem: Explain why missing data codes are important, and how to select the good missing data codes? +2 points

If an answer is longer than the given place for it, continue on the back of the sheet, but identify your answer properly.