

Assignment 5

NAME: Your Name

DUE DATE: April 25th, 6pm

Problem 1 (60 points)

In the folder Assignment 5, you will find the data set called `movie.rda`. This data set contains one movie in each row and a score (in the column) for each movie for the soundtrack, the picture, the director, and the originality.

- (a) (10 points) Check with elegant coding (cycle for or a function) which are the two variables containing the most NA values.
- (b) (25 points) Compute a random imputation for all the variables. Then compute the iterative regression imputation for the two variables containing the most NA (the ones discovered in point (a)).
- (c) (15 points) Compute the multivariate imputation by chained equations with the `mice` package. Plot (using `ggplot`) four histograms for the variable “Soundtrack”: 1- the variable without missing value, 2- the complete variable obtained with random imputation, 3- the complete variable obtained with the iterative regression imputation, and 4- the complete variable obtained with the `mice` package.
- (d) (10 points) Compute the gap statistic for the complete data set obtained with the `mice` package. Perform the k-means cluster with the optimum number of clusters.

Problem 2 (40 pts)

In the folder Assignment 5, you will find the data set called `popular2data`, containing data for 2000 pupils in 100 schools. The main outcome variable is the “pupil popularity”. The explanatory variables are pupil gender (boy = 0, girl = 1), pupil extraversion (10-point scale), and teacher experience in years.

- (a) (10 points) Perform a simple random intercept model, include all the covariate as fixed. Are the fixed covariates significant?
- (b) (15 points) Add to the previous model random slopes for both sex and extrav. What is the value of the error term (Variance) for the random slope of the variable sex? What does this value suggest to you?
- (c) (15 points) Omit from the previous model the random slope of sex. What is the value of the mean effect of extraversion? What is the value of the random effect of the slope of extraversion?