**Due Date: 13-Nov-2018**

PART 1: Cluster Analysis

The data labeled *unistudis.txt* contains the number of students from 13 different faculties in 72 universities who enrolled for a foreign language class. Perform a cluster analysis on this data to group the universities into clusters sharing similar student enrollment behavior(s). The description of the different faculties is provided tabulated below:

|  |  |
| --- | --- |
| Faculty symbol | Full Faculty Name |
| *FEB* | *Faculty of Economics & Business* |
| *FArch* | *Faculty of Architecture* |
| *FSS* | *Faculty of Social Sciences* |
| *FA* | *Faculty of Arts* |
| *FPES* | *Faculty of Psychology and Educational Sciences* |
| *FMeds* | *Faculty of Medicine* |
| *FLaw* | *Faculty of Law* |
| *FKRS* | *Faculty of Kinesiology and Rehabilitation Sciences* |
| *FBE* | *Faculty of Bioscience Engineering* |
| *FET* | *Faculty of Engineering Technology* |
| *FTRS* | *Faculty of Theology and Religious Studies* |
| *FCL* | *Faculty of Canon Law* |
| *FS* | *Faculty of Science* |

***UniID*** is a university identifier variable

PART 2: Discriminant Analysis

The dataset “*cycling\_data.txt*” contains socio-demographic and cycling background information for 521 Belgian respondents. Respondents were asked if they would consider using a bike in some specific circumstances. Information on the variables in the dataset is listed below.

* Choice: The respondent’s choice of using a bike for a specific route.
* Cycling\_Frequency: A variable on the scale of 1 to 5 that shows how often the respondent uses the bike. 1 indicates “almost never” and 5 indicates “almost every day”.
* Cycling\_skill: A variable on the scale of 1 to 5, which shows the cycling skill of the respondent. 1 indicates “no cycling skill” and 5 indicates “Advanced cyclist”.
* School\_by\_Bike: A binary variable with the value 1 if the respondent used the bike or is currently using the bike for his/her trips to school.
* Combine\_Bike: A binary variable with the value 1 if the respondent sometimes combine the bike with another means of transport. E.g. cycle to station and then take a train.
* Male: A binary variable with the value 1 for men.
* Age: respondent’s age
* Married: A binary variable with the value 1 if the respondent is married.
* Parent: A binary variable with the value 1 if the respondent has kids.
* Live\_in\_Flanders: A binary variable with the value 1 if the respondent lives in the Flanders and 0 if the respondent lives in Brussels or Wallonia.

1. Describe respondents who selected a bike versus those who did not select a bike?
2. Select an efficient set of discriminating variables? Discuss your result(s)