HEADS
PhD PROGRAMME IN HEALTH DATA SCIENCE

COMPUTATIONAL STATISTICS

1. Open the database Anesthesia-BD.csv

2. For the columns which contain numeric values, create a new summary table in which the rows are the mean, the standard deviation, and the median of each of the numeric columns.

3. Considering only the subjects who had CAB or Valve surgery, how many had adverse events? How many males and females in the two groups (having or not having adverse events)? (count and percentage)

4. Merge the two datasets (Anesthesia-BD.csv and Anesthesia-BD2.csv) by the subject identification number.

    a. The merge must be such that only the common subjects should be present in the final database (natural join).
    b. The merge must be such that if some subject is not in one of the databases, the subject should be in the database, and missing information must be not available (full outer join).

5. Consider the following statement: “For people without diabetes, the normal range for the hemoglobin A1c level is between 4% and 5.6%. Hemoglobin A1c levels between 5.7% and 6.4% mean you have a higher chance of getting diabetes. Levels of 6.5% or higher mean you have diabetes.” Create a new variable with three factors (normal, prediabetes, and diabetes), taking into consideration the values the A1c levels. Compare the results obtained with the variable Diabetes (assuming that 1 means to have diabetes and 0 no diabetes).

6. Create a table with the comparison between the groups having or not having adverse events for all the variables available in the combined database. Use the appropriate measures for each type of variable.