The Hang Seng University of Hong Kong
Department of Mathematics, Statistics and Insurance
AIN2740 Computing for Statistical Analysis

Exercise – R Dataset

R Code Question:

Follow step (i) to [illegible]nd print the dataset. Read Mark_gender1.txt and Mark_gender2.txt i[illegible]y 'C:\Users\your student ID', use dataset name as "mark_g1" and "ma[illegible] dataset should have 8 variables namely: sid, f_name, l_name, ca, mt, final, total, and grade. The mark_g2 dataset should have 4 variables namely: sid, gender, q1, and q2. Write comments in your code to indicate which part you are doing and then answer Questions (1) to (7).

(i) Merge dataset mark_g1 and mark_g2 by sid, make the combined dataset as mark1.

(ii) Convert variable grade and gender to categorical variables as factors for dataset mark1.

(iii) Omit incomplete cases of mark1 and rename the dataset as mark_c.

(iv) Subset observations in dataset mark_c where variable gender is F and grade A, sort the dataset in descending order of variable total, rename the dataset as mark_fAo.

(v) Subset observations in dataset mark_c where variable ca is over 70, then sort the dataset in ascending order of variable f_name, rename the dataset as mark_cao.

(vi) Use ifelse function to test the gender of students for dataset mark_c. Rename the result with a vector call gen_s showing "The student's gender is male" for M and "The student's gender is female" for F.

(vii) Create a for loop, use if function to test whether the variable total is greater than or equal to 80. Print the total if it is true.

Remember to save as your R Script file according to "StudentID_ChapterX" and upload. Read the submission guideline in Lecture Notes - Lesson 0 carefully before submitting to moodle.

(1) How many observation(s) are there in mark1 in part (i)?

(2) What is the 30^{th} element in the factor grade in part (ii)?

(3) What is the element of row 67 and column 7 of data frame mark_c in part (iii)?

(4) What is the row dimension of data frame mark_fAo?

(5) What is the element of row 14 and column 1 of data frame mark_cao?

(6) What is the 50^{th} element in the vector gen_s?

(7) What is the last element in the result in part (vii)?