HW Assignment (Linear Regression)

Note: Please hand in your assignment via email (hkkwon7@business.kaist.ac.kr) no later than 23:59 November 29th.

You will be required to hand in your answers (or R code/output) to all questions. You must be as explicit as possible regarding your answers and additional explanation is encouraged if necessary.

The dataset used in this section is an "App Consumption" dataset, and consists of 2 csv files (behavioral data and demographic data).

In the behavioral data, there are four variables: id, time, c. "id" refers to users' ids, "time" refers to days and "c" refers to the amount of time (seconds) an user spent on a specific app on a particular day. Hence, the first row would indicate that in day 1, user 1 spent 9 seconds on the app.

In the demographic data, there are 7 variables: id, sex1, sex2, job1, job2, job3, job4. "id" refers to users' ids, "sex1" and "sex2" are dummy variables indicating female and male respectively, and "job1", "job2", "job3", "job4" are dummy variables indicating student, businessman, housewife, others respectively. Hence, the first row would indicate that user 49 is a male with a job, "others".

Questions

a) Load the two csv files of behavioral data and demographic data

b) What is the key variable to join the two data?

c) Merge the behavioral data with demographic data

d) Save the merged data

e) Make dataset of the mean "c" for each id. Hence, the final data will have one row for each id with 8 variables (i.e., id, mean_c, sex1, sex2, job1, job2, job3, job4) (hint : *summaryBy*)

f) Is this dataset a panel data or a cross-sectional data? And why?

g) Run a regression analysis to examine the effect of gender and job on app usage. (hint : Dependent variable - "c", Independent variable - "sex1, sex2, job1, job2, job3, job4")

h) Screen capture the result

i) Interpret the results without considering significance level