

SMDE FIRST ASSIGNMENT (40% OF THE FINAL MARK)

FIRST QUESTION: VISUALISATION, CHI SQUARE AND T-TEST (25% OF THE FIRST ASSIGNMENT).

Download data set from the web site of Kaggle ([Technology Usage, Stress and Wellness](#)) and read the data set in R.

The data set consists of various variables on digital behavior (daily screen time, daily phone/laptop/tablet/tv/social media usage, work hours, entertainment hours, gaming hours), wellness related parameters (sleep duration, sleep quality, mood rating, stress level, physical activity, mental health score, using wellness applications, eating healthy, caffeine intake, weekly anxiety score, weekly depression score, mindfulness minutes per day) and some demographic characteristics of participants such as age, gender and location.

- a) Import data set to R assigning the type of each variable correctly. (0.5p)
- b) Classify variable age into 4 categories in a new variable named “age_cat” considering [the guide of generations](#) such that Baby Boomers (61-79 years), Gen X (45-60 years), Gen Y (29-44 years) and Gen Z (15-28). Name each level of the factor by corresponding name of generation. Similarly convert stress level variable to a factor named “stress_cat” which has two levels: “Low” (stress levels from 1 to 5) and “High” (including stress levels from 6 to 10). (1p)
- c) Analyze the association between variables “location” and “age generations (age_cat)” and then between “stress level (stress_cat)” and “age generations (age_cat)” by using proper method. Interpret your findings. (3p)
- d) Check the distribution of numerical variables by using a proper graph and test whether they follow a Normal distribution or not. Which of the variables follow a Normal distribution? Summarize your findings in the report. (1.5p)
- e) Test whether “mental health score” is affected from the stress level of participants. Apply correct method testing its assumptions and interpret the results. Enrich your findings by visualizing distribution of the variable “mental health score” for “Low” and “High” stress levels and using a boxplot. (4p)