

MTH1004 “Probability, Statistics and Data”

Term 1 Summative Coursework “Report 1”

Social media survey data analysis

- **Submission deadline: Friday 15 December 2023 by noon via ELE**

Assignment

Your task is to study, summarise and analyse survey data related to social media usage and mental health. You will produce numerical summaries and data visualisations, write brief summaries and discussions, and present your work in the form of a 1-page poster.

You may not collaborate with others on this coursework (including AI chatbots), and you may not use external resources without attribution. For detailed information on University policies regarding academic conduct and practice please refer to the online course on “Academic Honesty and Plagiarism” (via ELE) and Chapter 12 of the Teaching Quality Assurance Manual ([link](#)).

Data

The data file `survey.csv` used in this assignment contains a subset of responses from a survey on the relationship between social media usage and mental health, conducted in Bangladesh (Full data source: Islam et al. (2021) doi.org/10.1002/jcop.22675). The file `survey.csv` can be downloaded from the MTH1004 ELE page, under section “Assessments”. The file is in comma-separated values (csv) format, with one row per respondent and variables in each row separated by commas. Variable names are given in the first row of the file, with the following long explanation:

- **time:** time when the survey was taken
- **age:** age of the respondent at time of survey
- **gender:** gender of respondent
- **education:** highest level of education achieved
- **height:** height of respondent in metres
- **Q1:** Answer to question “How much time do you spend daily on social media?” (in hours)
- **Q2:** Answer to question “Which social media account do you use usually?”
- **Q3:** Answer to question “In the last 30 days, I felt nervous, anxious, or on edge.”
- **Q4:** Answer to question “In the last 30 days, I had trouble concentrating on things, such as reading the newspaper or watching television.”
- **Q5:** Answer to question “Over the past month, how would you rate your sleep quality overall?”

The following three parts of this assignment contain specific questions about social media usage that you should answer by analysing the data in `survey.csv`. More details about the required structure and format of your submission are given below.

To start, download the file `survey.csv` and read it into a data frame in R using the `read_csv` function from the `tidyverse` package. Use that data and the R commands learned in the lectures to answer the following questions.

Part 1: Introduction

1. How many people responded to the survey?
2. Over what time period did the survey run?
3. Which social media platforms do survey respondents use, and how many respondents use each platform?
4. Analyse the age distribution of respondents (median, lower and upper quartile), and comment on how representative the survey might be of the whole population of Bangladesh.
5. Use appropriate R commands to produce a small map that highlights the country outlines of Bangladesh, as shown in Figure 1.

Add your answers and visualisation to the poster. You can add additional summary statistics and visualisations to answer any of the questions.



Figure 1: Bangladesh

Part 2: Social media usage and sleep quality

1. Using suitable summary statistics, summarise the responses to the two questions related to sleep quality and daily social media usage.
2. Plot a suitable bar chart that visualises sleep quality among the different groups of daily social media usage, and briefly describe the bar chart.
3. Discuss whether you think social media usage might have an effect on sleep quality. Support your analysis by suitable summary statistics.

Add your bar chart, summary, and discussion to the poster.

Part 3: Social media usage and mental health

1. Using suitable data visualisations, summary measures and a brief discussion, answer **only one** of the following questions:
 - (i) Is there evidence that some social media platforms are more engaging or “addictive” than others?
 - (ii) Is there evidence that excessive social media usage has a negative effect on anxiety levels?
 - (iii) Is there evidence that excessive social media usage has a negative effect on concentration?

Add your visualisation, results and discussions to the poster.

Submission

Poster format

Your answers to the above questions should be submitted as a **one-page poster in pdf format**.

Your poster should be structured so as to contain an informative title, and three informative section headings corresponding to the three question parts. Each of the three sections should include a brief written summary of the results and answers to all questions of the corresponding part. Answers that do not appear on the poster will not be marked. Write in a clear and objective tone, and avoid slang or conversational language. You may include further information about the data or research that you think could make the poster more interesting. The text on the poster should be mostly written in full sentences. As a soft guideline, aim to write between 300 and 500 words in total.

The poster can be either in landscape or portrait orientation. The poster should be saved and submitted as a one page pdf file in A4 format. It must be possible to highlight and copy the text on the poster when viewed in a pdf viewer, that is, do not submit a pdf that just contains an image file of your poster. Do not submit your poster in any other file format than pdf (such as ppt, doc, png).

You can find more information on good formatting and writing on the Assessments page on ELE.

R code format

Submit all R code you used to produce your answers and figures as a **single R script file** (ending on `.R`). It should be possible to open and run the file in Rstudio without errors. The R code should be formatted for good readability, using appropriate spaces, indentation, line breaks, and code comments. Use R commands that you learned in the course as much as possible. If you use a function that was not covered in the course, add a comment on how or where you found it.

ELE submission

Please submit your two files (one .R file and one .pdf file) via the ELE submission point on the MTH1004 ELE page.

Marking guidelines

This summative coursework will be worth 15% of your final module mark.

Your poster will be marked against the following criteria, with some indicative questions that will be considered by the marker:

- *Accuracy 35%*: Are all reported answers correct? Do the visualisations display the correct data? Do the written descriptions match the visualisations? Does the summary provide a coherent description of the analyses and results? Do the conclusions logically follow from the data analysis results? Are the reported answers relevant to the questions?
- *Coding 35%*: Does the submitted R script run without errors? Does the code produce all results reported on the poster? Is the code well-documented, and formatted for good readability? Is there much code that is unrelated to the poster? Does the code contain many functions that were not taught in the course?
- *Writing style and poster design 30%* Is the written description of data and figures clear and concise? Does the text contain typos and grammatical errors? Does the poster look visually attractive, e.g. by using a consistent colour theme? Were font types, sizes, and colors chosen sensibly? Does the title attract attention?
- *Creativity bonus +10%*: A bonus of up to 10% can be awarded if the report contains additional interesting data analyses that were not explicitly asked in any of the questions. (But the total coursework mark cannot exceed 100%.)