University of Strathclyde
Glasgow

Department of Mathematics and Statistics

MM926 Survey Design and Analysis

Due Date: 8 December 2025   **16:00 (BST/GMT)**

Your project should be your own work. All submissions will be submitted to the plagiarism checker Turnitin to identify any similarity between student submissions and published online literature including websites. Students suspected of academic dishonesty, in terms of either plagiarism or collusion, will be investigated in accordance with the Student Discipline Procedure for Academic Misconduct.

Extension requests should be made via the module Myplace page. Extensions will be considered inline with the Policy and Procedure on extensions to coursework submission. This should also be referred to in the extension request.

Submissions made after the due date, without approved extension, will be subject to a deduction of marks in accordance with the Policy and Procedure for late submission of coursework.

This assessment is 2: "AI-Assisted Idea Generation and Structuring" on the generative AI usage scale.

CONTINUED ON NEXT PAGE

1. Project Information

This project is worth 65% of the overall mark for MM926.

In this project, you will conduct an analysis of survey data relating to the attitudes of undergraduate students towards social media. You will be required to submit a document containing the results from your analysis, a 3–5 minute video presentation summarising your analysis, along with slides for this presentation.

The 6-page limit stated is a strict limit. There will be a deduction in marks should you go over the page limit. This will be a 5% reduction for each page over the limit.

2. Project Background

The file "project_data.csv" contains data collected during a survey which aimed to assess the attitudes of undergraduate students towards social media in a particular university.

The survey data contains 20 variables:

- ID: a unique number used to identify a participant
- Age: age of participant
- Sex: male (0), female (1)
- Year: participant's year of study (1-5)
- Faculty: faculty that the participant belongs to (coded as 1, 2 or 3)
- Performance, Social, Mental_Health, Awareness, Connected, Resources, Expression, Community, Networking, Interactions, Time_Wasted, Concentration, Negative_Content, Data_Security, Inadequate: These are responses to LIKERT questions (next page), coded as:

strongly disagree (1), disagree (2), neither agree nor disagree (3); agree (4), strongly agree (5).

Survey participants were also given the opportunity to include a free text response about their overall attitude towards social media. Clear positive and negative comments from the responses are provided in the "positive comments" and "negative comments" text files.

CONTINUED ON NEXT PAGE

The statements which formed the LIKERT questions in the survey were as follows:

- Performance: Social media has had a negative effect on my academic performance.

- Social: Social media has enhanced my social interactions.

- Mental_Health: Social media has had a negative impact on my mental health.

- Awareness: Social media has a positive effect on social awareness.

- Connected: Social media helps me to stay connected with family and friends.

- Resources: Social media provides valuable academic resources.

- Expression: Social media provides me with a platform for self-expression.

- Community: Social media fosters a sense of community.

- Networking: Social media is useful for networking.

- Interactions: Social media reduces face-to-face interactions.

- Time_Wasted: I waste a lot of my time on social media.

- Concentration: Social media makes it difficult to concentrate on my studies.

- Negative_Content: Social media exposes me to negative content.

- Data_Security: I am concerned about my data security when using social media.

- Inadequate: I feel inadequate when I compare myself to others on social media.

CONTINUED ON NEXT PAGE

3. Project Tasks

1. Use the data provided to conduct the following analyses:

(a) Assess the internal consistency of the LIKERT questions. Would the removal of any questions improve the overall internal consistency?

(6 marks)

(b) Provide an appropriate visualisation of the responses to the LIKERT questions and use this to identify LIKERT questions that have (i) elicited a strong positive response from participants; (ii) elicited a strong negative response from participants; (iii) divided the opinion of participants.

(4 marks)

(c) Conduct an investigation into the missing data.

(19 marks)

In particular, you should:

- Provide an appropriate visualisation of the missing data.
- Test whether or not the data is missing completely at random.
- Impute missing data using regression imputation and predictive mean matching.
- Use appropriate plots to assess the two imputations for the variable with the most missing data. Comment on the results.
- Use Kolmogorov-Smirnov tests to assess the imputations for the variable with the most missing data. Comment on the results.
- Select an imputation method, with justification, and save the dataset that is created when this imputation method is used.

CONTINUED ON NEXT PAGE

1. (cont.)

(d) Using the imputed dataset saved in the previous question, perform unweighted logistic generalised linear regression to investigate whether sex or year is associated with a positive overall attitude towards social media, as indicated by an average value of 3.5 or more for the total score from the LIKERT questions. Provide an interpretation of your results.

(8 marks)

(e) Now suppose that participants were stratified based on the faculty that they belong to, and that there were a total of 2000 students in Faculty 1, 700 students in Faculty 2 and 300 students in Faculty 3. Use the imputed data from part (c) to construct a weighted logistic generalised linear model to investigate whether sex or year is associated with a positive overall attitude towards social media, as defined in (d). How do the results from the weighted regression compare to those from the unweighted regression?

(8 marks)

(f) Create a word cloud associated with positive qualitative responses and a word cloud associated with negative qualitative responses. Provide comments comparing the two.

(5 marks)

CONTINUED ON NEXT PAGE

2. Summarise your study in a 3–5 minute video presentation. You are not expected to discuss all of your analysis and findings but should instead pick out the aspects that you feel are most important and/or interesting. You should make slides for this presentation and record your video while sharing your screen so that both your slides and yourself are visible (e.g. using Zoom). Your presentation will be assessed on the following components:

• Content

(9 marks)

• Overall appearance of slides (including structure, spelling etc)

(2 marks)

• Overall presentation style (including clarity, confidence etc)

(2 marks)

• Timing

(2 marks)

4. Project Deliverables

For successful submission of the project, you should provide

- A Word or PDF document containing your results for Task 1 of the Project Tasks. This document should be no longer than 6 pages and should contain your answers to the questions as well as relevant R output. You MUST also include your R code in an appendix, but this appendix does not count towards the page limit.

- A 3–5 minute video presentation, as described in Task 2 of the Project Tasks.

- The slides that accompany your video presentation (in PDF or Powerpoint format).

END OF PAPER

LK

Total: 65 marks