Data Analysis Skills – Group Project 2

Objective: Create and present a group presentation summarising an analysis of data from an allocated dataset, with group members who did **not** attend the poster viva presenting the presentation. Analysis of the data must be worked on collaboratively using Quarto files stored and edited in GitHub.

Deadlines:

- Setup GitHub Repository – **12:00 Friday 8th March.**
- Select a time slot for group presentation – **17:00 Monday 11th March.**
- File submission – **09:00 Monday 18th March.**
- Contribution evaluations and Declaration of Originality – **09:00 Tuesday 19th March.**

Contribution to final grade: 25% (Marking Scheme detailed below)

Overview: Each group is tasked with analysing an allocated dataset which will require using a generalised linear model (GLM) in the formal analysis. Each group will create a GitHub repository to store, develop and share all code used throughout the analysis. The presentation will consist of a summary of the findings of the analysis.

Allocating tasks efficiently: In this project, there are two key components: the coding for the analysis using GitHub to store, review and approve changes, and creating and preparing the presentation. As such, you may wish to plan how you allocate these tasks. You are not all expected to update code and approve changes on GitHub nor will everyone in the group take part in the presentation, so it may work best if subsets of the group focus on different tasks.

For a group with 5 members, assuming group members 1 and 2 each attended a poster viva, one such workload model could be (**note – you do NOT have to follow this model!**):

- Group members 1,2 and 3 work on the coding on GitHub, checking each other's updates and any pull requests. Members 4 & 5 can also check any pull requests & commits to make sure these make sense.
- Group members 4 & 5 can begin to form the presentation, with member 3 providing the relevant code outputs. Members 1 & 2 can review the slides when required.
- Group members 3,4 & 5 take part in the presentation.

Each group must only use the MSTeams channel created by the lecturers for each group for communication throughout the project (please note that this excludes a MSTeams chat with all group members). All communication must be in English.

Questions for the lecturers should be posted to the Student Q&A Forum on the Moodle page.

Project Details

Stage 1. Analyse the allocated data set using GitHub to share and store code

The document "Datasets Descriptions" on the DAS Moodle page describes 20 datasets from which each group has been allocated one to work on for this assignment (the group number corresponds to the dataset allocated). Please ensure you are working with the correct dataset and are answering your assigned question of interest. The data sets are available to download from Moodle.

Analyse the data in a collaborative fashion in a file called **Group_##_Analysis.qmd** in GitHub in a repository named **DAS-Group-##** (where ## corresponds to the number of your group). Only one person will need to create the repository and upload the **dataset##.csv** file containing your allocated dataset, then add everyone else. Also add the course lecturers to your repository by **Friday 8th March at 12:00** using the GitHub usernames **craigalexander** and **mitchumbock**.

- **Ensure your repository is set to public.**
- When using your repository for analysis, make sure to leave detailed comments on any commits you make to the main branch (see Unit 4 of the Using Git course for effective strategies).
- Be sure to fork **at least one** piece of work out of the main branch. Assign members of your group to review the changes and then approve them using a pull request (see Units 5 and 6 from the Using Git course for details on how to do this).
- You will be asked to peer review relevant group members contributions to the repository, in terms of checking code, providing helpful commit messages, and approving pull requests with detailed comments.
- Your repository should follow general good practice with meaningful commit messages, a clear file structure, informative readme file and constructive review. You may also use additional features such as Issues for a checklist (see Unit 4 of Using Git course for some good practice tips).
- If you choose to save any relevant outputs on your repository, please put these in a folder with an appropriate title (e.g. Plots).
- The .qmd file which you share on Github must be able to render to a .pdf file which contains the commented code used in the analysis and appropriate output.

Stage 2. Produce a presentation on your analysis

Produce a set of slides which summarise the key findings in Stage 1. The slides should include the aims of the analysis, exploratory data analysis, statistical modelling and results, conclusions and future work/extensions. When creating your slides, consider the following:

- You may produce your slides using any format (Powerpoint, Quarto etc.) but it is best to use a format everyone in the group is comfortable using and can access. Powerpoint can be accessed/downloaded via Office365 on the University's webpage.
- The 'target audience' for this presentation are your fellow students on the MSc programme, i.e. you can assume some knowledge of statistical models and inference.
- Make sure your presentation has a logical flow, i.e. make sure there is a clear beginning/middle/end and a smooth flow between sections.
- No graphs or summaries or any other output used in the presentation can be 'copied and pasted' from another source, but all of the analysis presented must be reproducible from the Quarto and data files on GitHub.

Stage 3. Present an in-person presentation

Participate in an in-person presentation with each of the members of the group who **did not participate in a poster viva for Group Project 1** presenting**.**

- Your presentation should be up to 9 minutes long, with each person presenting for at least 2 minutes.
- Following the presentation, there will be 3 minutes of follow-up questions by the staff, with each member of the group required to answer at least one question.
- Two members of staff will grade individually, with grades being averaged at the component level (see marking scheme below) before being added to give a final mark.

Presentations will take place **at the times listed on the Course Information Sheet**.

Presentations will take place on-campus and the venues will be published at the start of the week beginning Monday 18th March. All members of all the groups that are scheduled to present during that presentation block are expected to attend. However, only the members of the group designated to present will be invited to present when their group's time comes.

A scheduler will be made available on Moodle which each group will use to select a time slot for their presentation and list the names of those presenting by **17:00 Monday 11th March**. After this deadline unallocated groups will be allocated a time by staff and the groups may be penalized for missing this deadline.

Submission instructions

You must decide on one member of each group to be responsible for submitting:

- Submitting **presentation slides** for each group in the file **Group_##_Presentation.pdf**
- Submitting the file **Group_##_Analysis.qmd** containing the analysis. (NB – you are **NOT** required to produce the presentation slides from the .qmd file)
- the file **Group_##_qmd.pdf** containing the .pdf document created by rendering the .qmd file
- the original file **dataset##.csv** containing the data explored in the analysis/presentation which can be read using the Group_##.qmd to reproduce the analysis.

The files must be submitted using the respective upload links in the "Week 9-11: Group Project 2" section on the Data Analysis Skills Moodle page. The deadline for uploading all the files is **09:00 Monday 18th March**.

Evaluation of individual contributions

In addition to submitting the slides and .qmd file per group, **each member** must complete the **Group Project 2 Contribution Evaluation** on Moodle. This will give you the opportunity to evaluate how well you feel you and your group members worked together. The evaluations by all group members may be used to assign different grades to individuals within the same group if there is evidence that individual members didn't contribute significantly.

The form will ask you to evaluate yourself and each of the group members on the following criteria:

- Collaboration
    - Listened to, valued, and supported the efforts and opinions of others.
    - Tried to keep people working well together.
- Preparedness
    - Had agreed work prepared to a sufficient standard for group meetings.
- Effort
    - Participated in and contributed meaningfully to group discussions.
    - Submitted high quality work
    - Was engaged and enthusiastic.
- Contribution
    - An active member of the group
    - Took their fair share of the workload
    - This may consist of, but is not limited to:
        - Interpreting the output obtained from the analysis.
        - Preparing and creating the final presentation.
        - Delivering the final presentation.

You will indicate whether you **Strongly Agree/ Agree/ Disagree/ Strongly disagree** that you and each group member demonstrated each of the qualities described above during the group project.

You will also be asked about your contributions to the GitHub page, with a series of Yes/No questions (**Please note – you do NOT have to answer Yes to all of these questions – be honest!)**

You will also be asked to:

- Detail your contribution to the project including any administrative roles you took on
- Detail the roles you and your fellow group members undertook in the project. Did these roles change as the project progressed?
- Reflect on the aspects of you and your group's teamwork that you thought were good and those that were not so good.
- Comment on how you found coding collaboratively on GitHub, e.g. did you find comments from others useful, were commits informative, etc.
- State whether there is anything different you have taken in your approach to working in a group compared to the previous group project.

Declaration of Originality

Together with the Group Project Contribution Evaluation, **each person** must make a **Declaration of Originality**. This declaration will be included at the end of the Contribution Form.

Each member of the group must complete the Evaluations and the Declaration 09:00 Tuesday 19th March.

Marking Scheme

The marking scheme for this assessment is split into 3 sections: Collaborative Coding & Analysis, Presentation Design & Content and Presentation. This assessment will be marked out of 60.

Collaborative Coding & Analysis [30 Marks]

The final version of the code stored in your group repository and uploaded to Moodle will be assessed. You will also be assessed on your use of GitHub to work in a collaborative environment.

Specifically:

- Appropriate exploration of the allocated dataset using multiple exploratory techniques has been conducted and any possible patterns/anomalies have been identified. [5 marks]
- Appropriate statistical methods have been correctly applied. [10 marks]
- .qmd file and its corresponding .pdf is tidy, reproducible, and **well commented** [10 Marks]
- **At least one** branch created to look at a section of analysis. This branch should be then merged into the master branch using a pull request. **At least one** member of the group should be assigned to check the code changes, comment and approve the pull request. [2 marks]
- Demonstrate good practice of using GitHub. For example, provide meaningful commit messages, review others' commits to the repo, keep repository organised and tidy, provide a meaningful README file. [3 marks]

Presentation Design & Content [15 Marks]

- Description of background to problem and questions of interest. [2 Marks]
- Overall visual appeal of the slides/readability. [2 Marks]
- Overall organisation and structure of slides is clear. [2 Marks]
- Clear and proper explanations of analysis. [3 Marks]
- Appropriate use of tables and/or figures. [2 Marks]
- Presentation of key results and validity of conclusions. [2 Marks]
- Discussion of future work/extensions [1 Mark]
- Appropriate level for target audience. [1 Mark]

Presentation Delivery [15 Marks]

- Clarity of speech. [3 marks]
- Appropriate pace and length of speech. [3 marks]
- Interaction with slides/visual presentation & audience. [3 marks]
- Answers to markers questions. [6 marks]