

Data Analysis Skills – Group Project 2

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

Deadlines:

- Setup GitHub Repository – 12:00 Tuesday 13th July
- File submission – 17:00 Monday 26th July
- Contribution evaluations and Declaration of Originality – 17:00 Monday 26th July.

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 and share all code used throughout the analysis. The presentation will consist of a summary of the findings of the analysis and will be recorded.

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 recording 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 appear in the recording, so it may **work best if subsets of the group focus on different tasks**.

For a group with 5 members, 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 others updates and merge requests. Members 4 & 5 can occasionally check any merge requests & commits to make sure these are sensible and 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 record the presentation.

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

Presentation Details

There are three stages to creating the presentation:

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

The document “Datasets Descriptions” on the DAS Moodle page describes 28 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.Rmd** in GitHub by creating a repository on GitHub named **DAS2021-Group-##** where **##** corresponds to the number of your group. Only one person will need to create the repository, then add everyone else. Please also add Craig and Mitchum to your repository by **Tuesday 13th July at 12:00** using the GitHub usernames **craigalexander** and **mitchumbock**.

- When using your repository for analysis, make sure to leave detailed comments on any commits you make to the main branch.
- 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.
- 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 merges with detailed comments.
- If you choose to save output on your repository, please put these in a folder with an appropriate title (e.g. Plots).

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, RMarkdown etc.) but it is best to use a format everyone in the group is comfortable using and can access. Powerpoint can be downloaded via Office365 using MyCampus.
- 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 R Markdown file(s) on GitHub.

Stage 3. Record an oral presentation

Record a group video presentation with each of the members of the group who **did not participate in the poster viva for group project 1** presenting in the recording.

- Recordings must be **between 7 and 8 minutes long**, recorded via software of your choosing in .mp4 format, with each person presenting for at least 2 minutes
- Zoom or Microsoft Teams are recommended for recording your presentation, with “How to...” information on Moodle.
- The presentation recording must show both the slides and the presenter that is speaking (i.e. the presenters’ cameras must be switched on).
- Save the recording with the file name **Group_##_Presentation.mp4**

Submission instructions

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

- Submitting **one recording** for each group in the file **Group_##_Presentation.mp4**
- Submitting the file **Group_##_Analysis.Rmd** containing the analysis. (NB – you are **NOT** required to produce the presentation slides or any other document from the .Rmd file)

The files must be submitted using the respective upload links in the “Week 9: Group Project 2” section on the Data Analysis Skills Moodle page.

The deadline for uploading your recorded presentation and .Rmd file is **Monday 26th July at 17:00**

Evaluation of individual contributions

In addition to submitting the one recording and .Rmd 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:

- Comment on what you think your contribution to the group project was
- Comment on how well you think your group worked together
- 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 **by 17:00 Monday 26th July**.

Marking Scheme

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

Collaborative Coding & Analysis [25 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]
- **At least one** branch created to look at a section of analysis. This branch should be then merged into the master branch. **At least one** member of the group should check the code changes, comment and approve the merge request. [5 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. [5 marks]

Presentation Design & Content [20 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. [4 Marks]
- Appropriate use of tables and/or figures. [2 Marks]
- Natural flow between sections. [2 Marks]
- Presentation of key results. [2 Marks]
- Validity of conclusions. [2 Marks]
- Discussion of future work/Extensions [1 mark]
- Appropriate level for target audience. [1 Mark]

Presentation Video [15 Marks]

- Clarity of speech. [4 marks]
- Appropriate pace of speech. [4 marks]
- Interaction with slides/visual presentation. [3 marks]
- Quality of recording. [4 marks]