Masters Programmes in Statistics and Data Analytics (session 2018/19)

Guidelines for Projects

1. Project overview

1.1. Eligibility

Details on student eligibility to progress to the project are available in the program handbook at

https://moodle.gla.ac.uk/pluginfile.php/1270127/mod_resource/content/2/MSc_handbook.pdf.

1.2. Project aim and intended learning outcomes.

The aim of the project is to teach (and assess) students in their ability competently to carry out an independent piece of (straightforward) statistical analysis. Specifically, students should be able to:

- Design an analysis plan to address the questions of interest posed about a data set.
- Implement that analysis in `R`, including, exploratory data analysis, model fitting, model checking and comparison.
- Draw appropriate conclusions from their analysis, including answering the questions of interest posed and identifying limitations and future work.
- Write up the results in the form of a report.
- Defend their analysis and conclusions in a mini-viva.

1.3. Workload and holidays

During the project students are expected to work full-time, e.g. roughly 30-35 hours per week, on their project throughout the 12 week summer project period. Our advice is that you are here in Glasgow working throughout the duration of the project, and that holidays are kept to a minimum. Remember that the project is worth 60 credits, a third of your entire degree.

1.4. Self proposed projects

Students proposing their own project will have to fill out a project description template, and this document can be found on the level 5 project Moodle page. Students have to have the data for the project available when they submit their project idea. The project team will vet the projects, and may suggest modifications or reject it if not suitable. If a student's self proposed project is rejected, then they will follow the standard process (outlined below) of choosing from the set of projects offered by staff. A drop-in session with the project coordinator will be offered for students needing help with creating their own project, the date and location of which is to be confirmed.

2. Timetable

The timetable for the MSc / MRes student projects is as follows.

- **26th February 2019** - Pizza lunch from 12 noon to 1pm in room 110 in the Mathematics and Statistics building to explain the projects process to the students.

- **23rd April 2019** - Deadline for project descriptions from staff and students.

- **20th May 2019** - Project descriptions are released to students, and the outcome of student proposed projects is given to the students involved.

- **31st May 2019** - Deadline for students to select their top 10 project preferences in ranked order.

- **10th June 2019 − 30th August 2019** - Project duration, specifically:

    - **10th June 2019** - Projects start.
    - **11th June 2019** - Pizza Lunch (12noon - 2pm Room 110).
    - **17th June − 21st June** - Week 2 project supervisory meeting.
    - **8th July − 19th July** - Week 5 and 6 project supervisory meeting.
    - **29th July − 9th August** - Week 8 and 9 interim assessment.
    - **30th August 2019** - Project submission deadline by 12noon.

Project allocation is final and no appeals are allowed.

3. Project supervision

All projects will be supervised collectively by a team of staff members, and because of holidays / conferences / etc there will not be a direct one-to-one staff-student link. The staff members on the project supervision team include:

- Craig Anderson
- Duncan Lee
- Vincent Macaulay
- Claire Miller
- Gary Napier
- Ruth O'Donnell.

Project supervision will consist of both drop-in sessions and limited scheduled one-to-one meetings as follows.

3.1. Weekly drop-in sessions

Drop-in help sessions for all students will be run on

- Tuesday mornings from 10am - 12am (rooms 110/117); and on
- Thursday afternoons 2pm-4pm.

Sessions on Thursdays are not run during weeks of the supervisory meetings, although the exception to this is the week of 24th June, when there will be a session on Thursday 27th June but not on Tuesday 25th June. A complete calendar can be found at the end of this document.

Each session will be staffed by 2 members of the project supervisory team and additional staff. These sessions will begin on Tuesday 11th June 2019 and the last one being on Thursday 22th August 2019. Students are welcome to come at any time during the session to ask questions.

3.2. Formal project meetings

The projects last 12 weeks starting on Monday 10th June, and the deadline for submission is Friday 30th August at noon. During the project period there will be the following formal project meetings.

- **Week 2. 17th June $- 21^{st}$ June**. Each student has a 30 minute supervisory meeting with one staff member to discuss their analysis plan. Guidance on writing a statistical analysis plan is available on the Project Moodle page.

- **Weeks 5 and 6. 8th July $- 19^{th}$ July**. Each student has a 20 minute supervisory meeting with one staff member to discuss their approach and the correctness of their analysis to date.

- **Weeks 8 and 9. 29th July $- 9^{th}$ August**. Each student has a 30 minute meeting with two staff members, which includes both formal assessment and supervisory elements.

    - **Assessment** - A short presentation (suggested maximum of 10 slides) on their project results so far, followed up with a viva which includes questions on all aspects of their project (including R code). This will last 20 minutes, a maximum of 10 minutes for the talk and a maximum of 10 minutes for the viva. Students will be asked to stop their presentations after 10 minutes.
    - **Supervisory meeting** - A short meeting for 10 minutes to discuss any further project problems / issues (this part is not graded).

Notes

- For each individual meeting a sign-up sheet on Moodle will be made available for students to select a time for this meeting.

- The formal sessions above will not necessarily be run by the same staff members each time, so the students will likely see different staff in the three weeks.

- Thus, students are expected to **upload their work that was discussed at each supervisory meeting (e.g. their analysis plan in week 2) to Moodle within 3 days of the session**, which will enable effective continuity of supervision between the 3 supervisory meetings run by different members of the supervisory team.

- You have to attend the meeting you have booked, if you miss them then you lose them. Good cause can of course be submitted in the usual way on MyCampus, if required.

- Students with Datalab projects are welcome to attend all of the above sessions, but are only required to attend the interim assessment in Weeks 8 and 9.

3.3. Skills week

The third week of the project period will include one-off lectures and workshops given by staff on core project competencies. The timetable is as follows:

- **25th June 2019, 10am-11:30am** - Scientific Writing - Group 1 (Room 109/110)

- **25th June 2019, 11:30am-1pm** - Scientific Writing - Group 2 (Room 109/110)

- **25th June 2019, 2pm-3pm** - Bibliography Writing (Room 116)

- **26th June 2019, 10am-11:30am** - Writing Maths - Group 1 (Room 109/110)

- **26th June 2019, 11:30am-1pm** - Writing Maths - Group 2 (Room 109/110)

- **27th June 2019, 10am-11:30am** - Presenting R Output - Group 1 (Room 117)

- **27th June 2019, 11:30am-1pm** - Presenting R Output - Group 2 (Room 117)

For the morning sessions (Scientific Writing, Writing Maths and Presenting R Output) the class will be split in two halves and you will attend only one of the two sessions. There is a

sign up sheet on Moodle for choosing which session you want to be in. On Tuesday 25th of June there will not be a drop-in session.

3.4. Other contact with supervisory staff

Projects are meant to be done independently, so normally, no extra contact with the supervisory team outside the ones listed above, will be allowed, in the interests of both staff workload and fairness to all students. This includes electronic communications. No emails in regards to any aspect of the project content will be answered by the supervisory team. Aspects of the project on which help is needed can only be brought to the weekly drop-in sessions or to one of the three formal supervisory meetings. These may include questions on basic `R` programming, standard statistical methods, interpreting results and report structure. However, draft project dissertations will not be read. Questions regarding administrative aspects of the project (e.g. "When is my viva?") can be addressed to the MSc Project coordinator (but not to members of the supervisory team).

3.5. Peer support

There are two sources of peer support throughout the project:

- The level M Moodle project page includes a **Peer Support Forum** where you may ask general questions about the projects to your peers. Questions may range from basic R programming questions (e.g. how do I delete a variable in a data frame) to admin ones (e.g. how do I sign up for the supervisory meeting). The forum is moderated by the project coordinator and will be closed down if misused.

- A peer review of a short draft (maximum of 5 pages) of your project will take place around week 6/7 (19th-26th July). Details about this are on the course Moodle page but it will be a great opportunity to have feedback on your progress from your peers.

Notes

Datalab students do not participate in the peer review mechanism due to disclosure issues.

3.6. Project coordinator

The MSc Project coordinator is Dr Manuele Leonelli, and all questions about the project process should be addressed to him at *stats-msc-project-coordinator@glasgow.ac.uk*.

3.7. Implementing project analyses

The default programming language used for the project is `R`. Students are allowed to use other programming languages if they are proficient with them, but this is likely to be very

rare and would need to be flagged up in advance of the project and vetted by the project supervisory team for feasibility. In this case students should be aware that the supervisory team may not be able to offer any help on coding.

4. Dissertation

The basic ingredients of a good dissertation are straightforward language and clear and well-annotated diagrams. The Introduction and Conclusions chapters should contain as little statistical jargon as possible. You are encouraged to include references to any sources consulted (including lecture notes, books, journal articles and web sites) and to cite all of them carefully in a References section towards the end of your dissertation. Your dissertation should be structured as appropriate but, in general, a rough outline of a well-structured dissertation might involve four sectioned chapters as follows. Note, these are suggested sections and not definite section headings!

- **Chapter 1 - Introduction to the problem**
  - Discussion of the context
  - Aims of the proposed research
  - Questions of interest
  - Description of the study and variables involved
- **Chapter 2 - Description of the methods**
  - Description of the statistical methods used
- **Chapter 3 - Standard analysis of the data**
  For each question of interest in turn:
  - Exploratory data analysis
  - Analyses and model checks
  - Interpretations of analyses
  - Conclusions for specific questions
- **Chapter 4 - Advanced analysis of the data**
  Use a more advanced method to answer the question(s) of interest which improves upon the simpler analysis in Chapter 3.
  - Analyses and model checks
  - Interpretations of analyses
  - Conclusions for specific questions
- **Chapter 5 - Conclusions and discussion**
  - Summary of conclusions to all questions of interest
  - Discussion of any limitations of data and/or analysis
  - Further analysis you did not have time to carry out
- **Appendix - R code** - You need to submit a separate `.R` file with your dissertation that contains the `R` code that produces the results (e.g. plots, tables, model fits, etc) in your dissertation.

The standard analysis should be achievable by everyone, and will be simpler methods that you have learned in your lecture classes. You can ask for help from the project team on this

analysis at the dedicated supervisory sessions. The advanced analysis may be an advanced method you have learned in your lecture classes or a method you have researched yourself, and should be researched completely independently. The aim with this chapter is for you to show independent research skills.

5. Project grading and submission

The project is assessed with two components:

- Dissertation
  - Main report excluding the advanced extra chapter - 50%,
  - Advanced extra chapter - 30%,
- Interim assessment (week 8) - 20%,

and details of each component are given below. Marking grids for these two components are available on the level 5 project Moodle page.

Dissertation
The project report can be written in any report writing software, such as R-markdown, MS Word/ LaTeX/LibreOffice etc. In all cases the R code should be provided as a separate .R file. The report should be a **maximum of 20 pages in 12 point font**, excluding title page and references. The only addition to this 20 page limit is the the R code, which should be provided as a separate .R file. We would expect that **around 5 of these 20 pages are dedicated to the advanced extra chapter.** The dissertations will be double marked by members of the project team.

Interim assessment
A 20 minute slot for each student with 2 members of the project supervisory team. During this session students will have to give a short presentation (suggested maximum of 10 slides) on their project results so far, followed up with a viva which includes questions on all aspects of their project (including R code). The 20 minute slot will include a maximum of 10 minutes for the talk and a maximum of 10 minutes for the viva.

Both the dissertation and interim assessment will be double marked by members of the project team on the 22 point scale.

6. Project submission

Dissertations must be be submitted electronically via Moodle on the level M Project Moodle page at

https://moodle.gla.ac.uk/course/view.php?id=4744

by 12 noon on Friday 30th August 2019.

7. Calendar for drop-in sessions

Below is a complete list of dates / times / locations for the drop in sessions. All sessions are in the Mathematics and Statistics building. For those sessions with 2 rooms listed, room 117 is the main room and 110 will only be used if needed due to a high volume of students. So please bring a laptop if you have one just in case.

DATE	TIME	Room
11th June 2019	10am-12am	110/117
13th June 2019	2pm-4pm	117
18th June 2019	10am-12am	110/117
27th June 2019	2pm-4pm	117
2nd July 2019	10am-12am	110/117
4th July 2019	2pm-4pm	117
9th July 2019	10am-12am	110/117
16th July 2019	10am-12am	110/117
23rd July 2019	10am-12am	110/117
25th July 2019	2pm-4pm	117
30th July 2019	10am-12am	110/117
6th August 2019	10am-12am	110/117
13th August 2019	10am-12am	110/117
15th August 2019	2pm-4pm	117
20th August 2019	10am-12am	110/117
22nd August 2019	2pm-4pm	117

8. Calendar for student study space

In addition to the library the following computing labs will be available for students to work on their projects during the summer.

- Lab 117 Mathematics and Statistics building - all summer.
- Lab 418 Boyd Orr building - Monday 10th June until Friday 21st June and Thursday 1st August until Friday 30th August.
- Lab 420 Boyd Orr building - Monday 24th June until Friday 30th August.