

CMEE Masters: Miniproject Assessment

Assignment Objectives: To address on a model-fitting problem using computational methods, and produce a written report, all in a coherent, reproducible, modular workflow under version control.

Student Name: Donal Burns

Overall Project Organization

You were missing your results directory, which may well cause you problems when I come to run your code! Use a `tt` `readme` or `.gitignore` file to make sure Git pushes ‘empty’ directories. Otherwise well organized, not clutter.

You had a `readme` file listing the key files and what they do in some detail – good! You should also have included packages /dependencies used (plus what they were used for and ideally version), and the R/Python versions used. Your instructions for running the project were explicit – excellent well done!

You could have put the writeup \LaTeX source files in a separate directory – this is what you should aim to do for your final dissertation.

Your organization was good, but the missing results directory is likely to cause problems. Your documentation is nicely explanatory/descriptive but missing a few key elements which prevents it from being outstanding. Overall, good.

The Code

Your choice of coding tools and number of packages used is broadly appropriate. You have used a sensible number of packages, however as a stylistic note you have imported entire modules and then gone on to import specific methods from those modules separately. I can see how it simplifies your code later on, but this is a slightly odd way of doing this. Check out the Pep8 style guide for Python (its conventions work fairly well in R too!).

Your code was clean and neat, and you have made reasonably logical use of R vs Python. You have commented well and made excellent use of docstrings (good job!). Your Python code is reasonably well modularized. Where you have modularized in your Python code, you have done an excellent job, giving each method a detailed docstrings. There will be a certain amount of unavoidable hard coding involved in a complex project such as this one, but in both your Python and R it would be worth thinking about ways you can make your code more compact, compartmentalized and general.

Running the project gave an error because you pass a data type that cannot be cast to `float64` (you pass a `dtype('O')` which is an object, which cannot be cast to `float64`). You also get an error when trying to write to your results directory because it is missing! In many programming languages file writing functions will not create a new directory whilst also writing a file, this is considered a safety/security feature generally, but means you must make sure a directory exists before trying to write to it (you can create directories explicitly, but not with file writing functions).

The data type casting error was fatal in that it prevented your subsequent scripts from running correctly or at all (due to missing data files!). Your L^AT_EX report was OK considering, the graphs were missing (because your scripts would not run) and you had a lot of warnings about colours printed to the console, also your bibtex threw many warnings (but this is fairly common). I cannot say whether or not your project ran in a timely manner, but for complex projects like this profiling is often useful, so consider it in the future!

Remember to write commands into your workflow which delete all previously generated files, so that everything is generated afresh for each run. Also add some commands which abort the workflow if the analysis encounters any errors!

I can see that had your scripts run according to plan you would have displayed progress messages to the terminal, which would have been clearly delineated. Good! An informed user is a happy user.

Your organization and documentation are reasonable, and your code is neat, clean and sensible. The errors are a shame, but if those were fixed this would have been a very very good project.

Overall still a very good effort.

The Report

Overall, a nice study. Reaching a good level of technical sophistication. Falling short in a few technical aspects (see comments further below), writing style, and presentation quality.

In your main project report, to the extent possible, consider using a results-focused title for your main dissertation — basically, the main finding(s) of the paper can be indicated in the title. Obviously, if you have a lot of findings, you can't do that. But then, it means you have too many 'storylines' in your study.

In your main project report, make sure your Introduction is more explicit about your objectives and hypotheses, and provided clear answers to them at the end, in the Discussion.

Have a look at this for further guidelines on figures for your main dissertation: http://abacus.bates.edu/~ganderso/biology/resources/writing/HTW_Guide_Table-Figures_9-30-08.pdf.

(Specific feedback is in the attached pdf, and we can also discuss more aspects of your write-up in our 1:1 feedback meeting)

Overall Assessment

Marks for the project and computational workflow: 75

Marks for the Report: 65

Overall Marks (*50:50 Computing:Report weighting*): 70

Signed: Samraat Pawar

June 2, 2020

Notes on Assessment :

- This written feedback will be discussed in a 1:1 session scheduled after this assessment has been given to you.
- The coursework marking criteria (included in this feedback at bottom) were used for both the computing and report components of the Miniproject Assessment. *In contrast*, Your final dissertation project marks are going to be based pretty much exclusively on the written report and viva (not code). Expect your final dissertation report to be marked more stringently, using the dissertation marking criteria (also included in this report).
- If there were technical errors made in the model fitting and selection, the points have been deducted from the report (not the computing) component's marks.
- In many cases, the marker would have contrasted what you have done with what you should do in your actual dissertation. *This does not mean that you were penalized* — one of the main goals of the miniproject is to provide feedback useful for your main dissertation. However, there may be cases where what you have done is just really bad practise (for example missing line numbers or abstract), irrespective of whether it is a mini- or main-project report – you will be penalized in that case.
- The markers for this assessment are playing the role of somebody trying to understand and use your project organization and workflow from scratch. So it will seem like the feedback is particularly pedantic in places — please take it in the right spirit!
- Ultimately, keep in mind that this mini-project was partly an exercise in reproducible workflow development — you may need to trade-off some computational elegance (but hopefully not reproducibility!), such as having everything run with one `run_project` command, in favor of a good written report — that's what matters most in the end. In this context, the main thing to keep in mind is that one or both of your markers will likely not be particularly quantitative, so you will need keep the explanations simple (but not patronizingly so!). In general, this advice holds while writing papers for more general (not narrow-subject focused) journals as well — keep it succinct and simple. Therefore, please also consider the report component mark separately from the computing component mark.

MARKING CRITERIA for EXAMS and ESSAYS and COURSEWORK

The following criteria are the basis on which the Department assesses both exam answers and coursework.

Literal Grade	Criteria (Problem type answers are marked on a semi-absolute scale)
A*	Exceptional Answer is an exceptionally well presented exposition of the subject, showing: (i) command of the relevant concepts and facts, (ii) a high critical or analytical ability**, (iii) originality, and (iv) evidence of substantial outside reading (where applicable). Numeric marks available 100, 95, 90, 85.
A	Excellent Answer is a very well presented exposition of the subject, showing many of the above features, but falling short in one or two of them. Numeric marks available 80, 76, 72.
B	Very Good to Good Answer (i) shows a clear grasp of the relevant concepts and facts, (ii) gives an accurate account of the relevant taught material (<i>as exemplified in the model answer</i>), and (iii) shows evidence of some outside reading or of critical or analytical ability**. Numeric marks available 68, 65, 62.
C	Adequate Answer: (i) shows a grasp of the basic concepts and facts, (ii) gives a mainly accurate account of at least half of the relevant taught material (<i>as exemplified in the model answer</i>), and (iii) does not go beyond that, or goes beyond that but is marred by significant errors. Numeric marks available 58, 55, 52.
F	Unsatisfactory Answer: 1.shows only a weak grasp of the basic concepts and facts, and is marred by major errors or brevity; numeric marks available 48, 45, 42; 2.shows a confused understanding of the question; is too inaccurate, too irrelevant, or too brief to indicate more than a vague understanding of the question; 35, 30, 25; 3.includes at most one to four sentences or facts that are correct and relevant to the question; numeric marks available 20, 15, 10, 5; 4.contains nothing correct that is relevant to the question; numeric mark 0.

** *Analytical* = assessing a hypothesis or statement by breaking it down into its elements and examining their inter-relationships and contribution to the whole; cf. *Critical* = judging a hypothesis or conclusion by examining the validity of the evidence adduced for it.

**MSc & MRes PROJECT ASSESSMENT – THESIS (OR MID-PROJECT REPORT)
MARKING CRITERIA**

Literal Grade	% Grade	Criteria (Please give leeway if it is a mid-project report)
A*	100 95 90 85	Exceptional. Work is of a publishable standard**. It is an exceptionally well presented exposition of the project, showing: (i) command of the relevant concepts and facts, (ii) a high level of analysis, (iii) originality in thought and experimental or modelling design, and (iv) mastery of the relevant literature.
A	80 76 72	Excellent. Thesis is written to a publishable standard** with minor revision. It is a very well presented exposition of the project, showing most of the above features, but falling short in one of them.
B	68 65 62	Very Good to Good. Thesis contains potentially publishable material**, but needs revision of the text and further research. It is otherwise a well presented exposition of the project, showing: (i) a clear grasp of the relevant concepts and facts, (ii) appropriate, though not highly sophisticated analysis, and (iii) a sound knowledge of the relevant literature.
C	58 55 52	Adequate. Thesis is not written to a publishable standard and requires major revision and substantially more research. It is an adequately presented exposition of the project, showing: (i) a grasp of the basic concepts and facts, (ii) an adequate use of statistics in its analyses, and (iii) sufficient knowledge of the relevant literature to set its results in a scientific context.
D	48 45 42	Unsatisfactory. Thesis is an incomplete presentation of the project and is marred by major errors or gaps, missing analysis, lack of references, misconceptions, excessive brevity, etc, at most showing a weak grasp of the basic concepts and facts.
	35 30 25	Thesis as above, but presentation extremely poor and overall impression indicates a very weak grasp of the basic concepts and facts.
	20 15 10	Thesis as above, and in addition no real attempt to analyse data or present results in a scientific manner.
	5	Thesis as above but incomplete and lacking understanding in all areas.
	0	Thesis not produced.

** This publishability implies that the data or theory is *per se* worth publishing.