Data Science for Health and Biomedical Sciences
BIME10076

Honours Elective
Deanery of Biomedical Sciences
University of Edinburgh
2024-25

Contents

If you require this document or any of the internal University of Edinburgh online resources mentioned in this document in an alternative format, please contact the Course Administrator <u>Stewart Smith</u> or <u>BMTO@ed.ac.uk</u>.

The contents of this handbook apply for the session year stipulated. The University may make changes to this course for future sessions.

Teaching Staff

Course Organiser	Dr Kasia Banas
	Kasia.Banas@ed.ac.uk
Course Staff	Keith Douglas
	Kdougla3@ed.ac.uk
	Dr. Karim Rivera-Lares
	lrivera@exseed.ed.ac.uk
Course Administrator	Stewart Smith
	stewart3.smith@ed.ac.uk

For queries, please contact the Course Administrator in the first instance:

BMTO, Medical School
Teviot Place
Edinburgh, EH8 9AG
stewart3.smith@ed.ac.uk

External Examiner:

Dr. Nicola Rennie (Lancaster University)

Please note that students should not make direct contact with the External Examiner. If you have any queries with the assessment process, contact the Course Administrator.

Course title: Data Science for Health and Biomedical Sciences
Course code: BIME10076
SCQF credits: 20 Credits
Credit level: SCQF Level 10 (Year 4 Undergraduate)
DRPS: http://www.drps.ed.ac.uk/24-25/dpt/cxbime10076.htm
Main teaching location: Central campus

Course Aims and Objectives

Data science is revolutionising how medicine is understood, how biomedical research is conducted and how healthcare is delivered. Despite the widely recognised opportunities that data can bring to biomedicine and healthcare, there is a shortage of data skills in the healthcare sector.

The course aims to provide a broad introduction to data science in health and biomedical sciences, covering key concepts and principles, data analysis skills and implications of working with biomedical and healthcare data. Key topics in the course include types of human health data; computational methods (e.g. process modelling and machine learning); data wrangling, analysis and reporting using the R programming language; legal considerations and bias in health data. This course is delivered in a flipped classroom format: it is based around short, pre-recorded videos, which are complemented with readings and self-guided programming tasks. There will also be weekly in-person tutorials and drop-in sessions, which provide an opportunity for further improvement of programming practice and discussion of the core concepts. Weekly quizzes will provide students with valuable formative feedback.

Learning Outcomes

By the end of this programme, students will be able to:
- Apply a range of specialised data science techniques to different medical and healthcare scenarios.
- Analyse health and biomedical data with the use of the R programming language, including summarisation, visualisation and interpretation.
- Critically examine the ethical, societal and regulatory principles and implications of data science in health.
- Explain and critically discuss key concepts, principles and methods of data science in health.

Graduate Attributes

- Digital literacy and numeracy, including using advanced data analysis tools to support their research and enquiry.
- Critical and analytical thinking, including applying critical analysis, synthesis and evaluation to key approaches and development in the subject.
- Communication, including communicating complex ideas and arguments to a range of audiences with different levels of knowledge/expertise.
- Personal and intellectual autonomy, including planning organising work, time management and taking responsibility for own work.
- Employability, including key data science skills that are in high demand among employers globally

Timetable

Course activities will take place on Thursdays and Fridays from 14:00 – 17:00

Week	Date & Time	Activity (lecture/ presentation/ seminar/tutorial)	Title/Topic	Tutor	Location
1	Thurs 19 Sep 14:10	Workshop	Introduction to data science for health and biomedical sciences	KD, KRL	G10 - Drummond Library - Old Surgeons Hall
	Fri 20 Sep 14:10	Workshop		KD, KRL	LG.08 - 40 George Square Lower Teaching Hub
2	Thurs 26 Sep 14:10	Workshop	Basics of R and the Tidyverse	KD, KRL	G10 - Drummond Library - Old Surgeons Hall
	Fri 27 Sep 14:10	Workshop		KD, KRL	LG.08 - 40 George Square Lower Teaching Hub
3	Thurs 3 Oct 14:10	Workshop	Machine learning techniques	KD, KRL	G10 - Drummond Library - Old Surgeons Hall
	Fri 4 Oct 14:10	Workshop		KD, KRL	LG.08 - 40 George Square Lower Teaching Hub
4	Thurs 10 Oct 14:10	Workshop	Reporting in R	KD, KRL	G10 - Drummond Library - Old Surgeons Hall
	Fri 11 Oct 14:10	Workshop		KD, KRL	LG.08 - 40 George Square Lower Teaching Hub
5	Thurs 17 Oct 14:10	Workshop	Introduction to Geospatial Data & Functions	KD, KRL	G10 - Drummond Library - Old Surgeons Hall
	Fri 18 Oct 14:10	Workshop		KD, KRL	LG.08 - 40 George Square Lower Teaching Hub
6	Thurs 24 Oct 14:10	Workshop	Analysing and presenting data in R	KD, KRL	G10 - Drummond Library - Old Surgeons Hall
	Fri 25 Oct 14:10	Workshop		KD, KRL	LG.08 - 40 George Square Lower Teaching Hub
7	Thurs 31 Oct 14:10	Workshop	Process improvement. Prescriptions and population	KD, KRL	G10 - Drummond Library - Old Surgeons Hall

	Fri 1 Nov 14:10	Workshop		KD, KRL	LG.08 - 40 George Square Lower Teaching Hub
8	Thurs 7 Nov 14:10	Seminar	Queer data workshop	Dr. Kevin Guyan	G10 - Drummond Library - Old Surgeons Hall
	Fri 8 Nov 14:10	Seminar	Data ethics seminar		LG.08 - 40 George Square Lower Teaching Hub
9	Thurs 14 Nov 14:10	Seminar	Visibility of What? Sociological and historical discussions of how we make and use data in clinical & biomedical knowledge production	Dr. Max Perry	G10 - Drummond Library - Old Surgeons Hall
	Fri 15 Nov 14:10	Seminar	Data ethics seminar		LG.08 - 40 George Square Lower Teaching Hub
	Fri 15 Nov 13:00	*Formative assessment due - R programming assignment section*			
10	Thurs 21 Nov 14:10	Workshop	Working with genomic data	KD, KRL, SMP	G10 - Drummond Library - Old Surgeons Hall
	Fri 22 Nov 14:10	Workshop		KD, KRL	LG.08 - 40 George Square Lower Teaching Hub
11	*Mon 25 Nov 13:00*	*Summative assessment due - R programming assignment*			
	Thurs 28 Nov 14:10	Space to sit and write	Dedicated time to work on your assessment – individual appointments available to discuss	KD, KRL	G10 - Drummond Library - Old Surgeons Hall
	Fri 29 Nov 14:10	Drop-ins		KD, KRL	LG.08 - 40 George Square Lower Teaching Hub
12	*Mon 2 Dec 13:00*	*Summative assessment due – Essay*			

Staff

KB   Kasia Banas
KD   Keith Douglas
KRL   Karim Rivera-Lares
SMP   Sophie Marion de Proce

Location

Thursdays: G10 - Drummond Library - Old Surgeons Hall
Fridays: LG.08 - 40 George Square Lower Teaching Hub

Recommended Reading

There is no compulsory course text.

Recommended materials include:

[R for Health Data Science](#) by Ewen Harrison and Riinu Pius
[R for Data Science](#) by Hadley Wickham and Garrett Grolemund
[Value sensitive design: Shaping technology with moral imagination](#) by Batya Friedman and David G. Hendry

Expectations

This course is taught in a flipped classroom format. This means that you will be given a set of readings and video materials each week, and you will be expected to engage with them in preparation for the live workshop. There is a self-study session timetabled each week, which you can use for this preparation work, but you're welcome to do the preparation at any time that is convenient, as long as you come to the workshop prepared. You will not benefit much from the workshops unless you do the preparation work beforehand.

All students will be expected to attend every teaching session, and if they are unable to do so they should send an e-mail giving the reason for their absence to the course organiser. As this is an integrated course, it is necessary to attend the whole course to allow you to draw upon material from several weeks to give an excellent answer in the assessments.

Code of Conduct

A lengthy code of conduct would be an overkill, so let me sum it up in three words:

Feedback

Students will be invited to complete a weekly formative quiz, and will be provided feedback on this.

Students will also be given an opportunity to submit a draft of one section of the programming assessment and will receive tutor feedback on this work.

Finally, the seminars will be structured to allow students to ask questions, and gain feedback on the data interpretation skills that they will develop throughout the course.

Students will be invited to provide feedback on the course through a mid-course questionnaire and end of course questionnaire.

Assessment

Assessment of this course is 100% in-course assessment (ICA).

ICA consists of:
1. R programming assignment (70%)
2. Essay (30%)

To pass the course you must achieve a mark of 40% or greater in BOTH components of assessments.

1. R Programming Assignment

This component of assessment is worth 70% of the overall course mark.

You will be asked to analyse a specific dataset and answer policy-relevant questions based on your analysis. This assessment will provide a balance of specific instruction (the dataset and questions will be provided by the teaching team) and autonomy (you will be able to decide how they would like to approach the question and will have the opportunity to focus on specific aspects of the data).

You will be able to submit a one-page section (preferably data visualisation with short commentary) by **Friday 15th November 2024 @ 13:00** for formative feedback.

You will be required to submit a formatted report, including your findings, code and data visualisations by **Monday 25th November 2024 @ 13:00**.

The maximum length for reports is 10 pages. This includes all sections of the report – text, code, figures and tables. Reports that exceed the 10-page limit (there is <u>NO</u> buffer limit) will be subject to a 10% penalty of the maximum mark.

Further details of the assessment can be found on Learn in the Assessment folder.

2. Essay

This component of assessment is worth 30% of the overall course mark.

You will be asked to explore some of the ethical challenges arising throughout the health data lifecycle, in the form of an academic essay. You will be provided with examples of problematic data-driven technologies, and will be asked to choose one, analyse it and recommend improvements.

Essay topics can be found on Learn in the Assessment folder.

The deadline to submit your final essay is **2nd December 2024 @ 13:00.**

The maximum word count for essays is 1200 words. The word limit for the essay does not include figure legends, tables or the final reference list, but does include in-text citations, headings and subheadings. Please include a final word count on the essay cover sheet. Essays that exceed the 1200 word limit (there is <u>NO</u> buffer limit) will be subject to a 10% penalty of the maximum mark.

Word counts will be checked by opening the essay file in MS Word and removing all sections not included in the final word limit. The word count then specified in MS Word will be taken as the final word count.

All in course assessments should be submitted to the drop box on Learn.

Work submitted after the deadline will be subject to penalties (see page 8). All written work will be submitted to a plagiarism check.

Summary of Deadlines

Assessment	Formative/ Summative	Submission Deadline	Feedback Return
R programming assignment section	Formative	13:00 Friday 15th November 2024	Monday 18h November 2024
R programming assignment	Summative	13:00 Monday 25th November 2024	Monday 6th January 2025
Essay	Summative	13:00 Monday 2nd December 2024	Monday 6th January 2025

Late Submission Penalties

1.1   Word Limits and Word Counts
Penalties for exceeding the word limit

Where an absolute word limit is specified in assessment documentation the following penalties will be applied if the word limit is exceeded.

- Exceeding word count by ≤ 10%: 10% deduction of the maximum mark or a reduction to the minimum pass mark for the assessment, whichever results in the higher mark.
- Exceeding word count by >10% and ≤ 20%: 20% deduction of the maximum mark or a reduction to the minimum pass mark for the assessment, whichever results in the higher mark.
- Exceeding word count by > 20%: mark reduced to the minimum pass mark for the assessment.

(b)   What is and is not included in the word limit.
- When evaluating the word count it **does not** include the assessment title, table of contents, references, text in figures, text in tables, table and figure titles, table and figure legends, a list of abbreviations or any acknowledgements.
- The word limit **does include** in-text citations (where Harvard referencing is required), headings and sub-headings.
- The use of appendices is not an allowable mechanism for exceeding the maximum word limit. Therefore, unless otherwise stated in the assessment guidance, material included in an appendix will be regarded as non-assessable and will not be considered by the markers/examiners when arriving at a final mark.

(c)   Additional Guidance: Figures and Table With regard to words in figure and / or diagram legends, a figure legend is primarily descriptive of the contents of the figure. The legend must not be used to introduce new

- data or information not referred to in the main text of the assessment. Tables should be used to summarise several pieces of information in a succinct and easily approachable manner. As such, figure legends that go beyond the description of figure contents and unnecessarily extensive tables will not be regarded favourably by markers.

Attempts to deliberately circumvent the word limit through, for example, inappropriate hyphenation, including blocks of text as picture images, etc. will be regarded as potential academic misconduct and referred to the Deanery Academic Misconduct Officer.

(d)   How the policy will be applied.
- Students will be required to indicate the total word count on the assessment cover sheet. Where a word limit is specified, word counts will be checked by the course administrator by opening submissions in MS Word and removing all sections not included in the final word limit. The word count then specified in MS Word will be taken as the final word count.

NOTE: If you experience problems uploading your document, do not panic; simply send the file directly to the Course Administrator (stewart3.smith@ed.ac.uk). The sent date and time on your email will be taken as the time of submission.

Note that any submissions made after the return of feedback dates will not be marked.

Course Work Extensions and Exceptional Circumstances

For students who need to request an extension for in-course assessments, or need to apply for exceptional circumstances, please refer to the 'Help and Support' section on Learn for further details. Note that the maximum period for eligible course assessments is 3 days.

Help and Support

Links to University Systems, BMTO Polices & Guidance and Central University Support can be found in the Help and Support section of Learn.

Assessment Criteria: In-course Essay

Class	Mark	Description
Fail	0-39	- Serious misunderstandings / errors throughout or mostly irrelevant material - Presentation, style and grammar very poor.
III	40-49	- Shortcomings in knowledge and understanding of topic - Little evidence of using original papers and mainly descriptive - Significant misunderstandings, errors or omissions - Poor presentation

IIii	50-59	• Good knowledge and understanding of the topic
		• Reference to original papers
		• Clarity of style
		• Minor misunderstandings or errors
		• Reasonable presentation
IIi	**60-69**	• Directly addresses the subject with good discussions
		• Very good knowledge and understanding of the topic
		• Logical structure with valid conclusions
		• Relates different pieces of information together and gives relevant examples
		• Evidence of a critical grasp of the topic
		• High standard of presentation including appropriate literature citations
I	**70-84**	• Critical judgement in selecting and evaluating relevant material
		• Evidence of wide independent reading and investigation
		• In-depth knowledge and understanding of the topic
		• Complete and appropriate valid conclusions of presented citations
		• Very high standard of style and presentation
	85-100	• Exceptional work and insight with original ideas
		• Synthesis of material including theoretical ideas, logically developed and incorporating evidence for statements made
		• Faultless presentation

Assessment Criteria: R programming assignment

[Click here](#) to see the assessment criteria