MODULE SPECIFICATION

Academic Year (student cohort covered by specification)	2026-27
Module Code	2489
Module Title	Statistics for Health Data Science
Module Organiser(s)	Elizabeth Williamson & Thiago Cerqueira Silva
Faculty	Epidemiology & Population Health
FHEQ Level	Level 7
Credit Value	**CATS:** 15 **CATS:** 7.5
Term of Delivery	Term 1
Mode of Delivery	For 2026-27 this module will be delivered by predominantly face-to-face teaching modes. Where specific teaching methods (lectures, seminars, discussion groups) are noted in this module specification these will be delivered by predominantly face-to-face sessions. There will be a combination of live and interactive activities (synchronous learning) as well as recorded or self-directed study (asynchronous learning)
Mode of Study	Full-time
Language of Study	English
Pre-Requisites	None, over and above the pre-requisites for the programme MSc Health Data Science
Accreditation by Professional Statutory and Regulatory Body	None
Module Cap (indicative number of students)	As per the number of student registrations on the MSc in Health Data Science.
Target Audience	This is a compulsory module for the programme MSc Health Data Science
Module Description	This module provides an introduction to the key statistical concepts and methods for health data science. Topics covered include probability, initial data description and exploration, statistical inference, regression, and Bayesian analysis. These topics provide the framework needed for

	subsequent modules. The module places a focus on learning through practical examples and incorporates directed learning, lectures, group discussion, and computer practical exercises.
Duration	18 x 0.5 day sessions over 10 weeks.
Timetabling slot	Term 1
Last Revised (e.g. year changes approved)	June 2023

Programme(s) This module is linked to the following programme(s)	**Status**
MSc Health Data Science	Compulsory

Module Aim and Intended Learning Outcomes

Overall aim of the module
The overall module aim is to: • introduce the motivation and critical thinking towards solving a question in health science through interrogation of data and drawing conclusions from evidence; and • introduce the principles of probability, regression modelling and statistical inference within frequentist and Bayesian settings.

Module Intended Learning Outcomes
Upon successful completion of the module a student will be able to: 1. evaluate the application of different probability distributions to model health data (including Poisson, Binomial and Normal); 2. critically analyse frameworks for frequentist and Bayesian inference and evaluate their strengths, limitations and differences; 3. examine the concepts of sampling variability, estimators, bias, confidence intervals and credible intervals; 4. examine the theoretical basis of linear regression and generalized linear models; 5. assess the application of regression modelling to address specific health data science questions; 6. critically evaluate strengths and limitations of different statistical methods, including regression models, within a health data science project; 7. draw conclusions from the results of a data analysis and justify those conclusions, appropriately acknowledging uncertainty in the results.

Indicative Syllabus

Session Content
The module is expected to cover the following topics: • Exploratory data analysis and the “problem solving” cycle • Conditional probability, Bayes theorem, binary/discrete distributions • Distributions for continuous variables • Sampling distributions and the central limit theorem • Likelihoods and maximum likelihood estimation • Frequentist inference • Bayesian inference • Regression modelling, linear models and extensions • Generalized linear models

Teaching and Learning

Notional Learning Hours

Type of Learning Time	Number of Hours	Expressed as Percentage (%)
Contact time	46.5	31
Directed self-study	43.5	29
Self-directed learning	34.5	23
Assessment, review and revision	25.5	17
Total	**150**	**100**

Student contact time refers to the tutor-mediated time allocated to teaching, provision of guidance and feedback to students. This time includes activities that take place in face-to-face contexts such as lectures, seminars, demonstrations, tutorials, supervised laboratory workshops, practical classes, project supervision as well as where tutors are available for one-to-one discussions and interaction by email.

The division of notional learning hours listed above is indicative and is designed to inform students as to the relative split between interactive and self-directed study.

Teaching and Learning Strategy
Each session will consist of an interactive session between students and lecturer, either as a lecture, seminar or practical. Students will be expected to read directed material ahead of the lectures/seminars and the learning outcomes will be consolidated during this time. This module will use a mixture of teaching techniques, including traditional lectures and

Teaching and Learning Strategy
practicals and student-centered discussions. Practical sessions will be provided that consist of problem-based learning exercises to further consolidate the learning outcomes into useable skills when facing a health data question. The practicals will include computer exercises and group discussion. Formative assessment will include an assignment where students are given a problem to solve using data.

Assessment

Assessment Strategy
The assessment for this module has been designed to measure student learning against the module intended learning outcomes (ILOs) as listed above. Formative assessment methods are used to measure students' progress. The grade for summative assessment(s) only will go towards the overall award GPA. Assessment will consist of formative practicals, a formative written assignment and an assessed exam at the end of the module. The formative written assignment will involve writing a short report on a statistical analysis and is designed to measure the students' progress. There is a dedicated session (session 7) to provide further information and time to complete the assignment. The main summative assessment will incorporate a mixture of multiple choice questions and traditional exam questions. The exam will be of length 2 hours and questions will be mapped to the intended learning outcomes through questions specific to each outcome.

Summative Assessment

Assessment Type	Assessment Length (i.e. Word Count, Length of presentation in minutes)	Weighting (%)	Intended Module Learning Outcomes Tested
Timed Test (in-module test e.g. MCQ)	2 hours	100	1-7

Resitting assessment
Resits will accord with Chapter 8a of the LSHTM Academic Manual. The resit will consist of a new timed test covering the ILOs.

Resources

Indicative reading list

Kirkwood B. R. and Sterne J. A. C. Essential Medical Statistics, Wiley-Blackwell, 2nd Edition, 2003. (available at the LSHTM online library) ProQuest Ebook Central, https://ebookcentral.proquest.com/lib/lshtmuk/detail.action?docID=624728.

Belle, Gerald Van, et al. Biostatistics : A Methodology For the Health Sciences, John Wiley & Sons, Incorporated, 2004. (available at the LSHTM online library) ProQuest Ebook Central, https://ebookcentral.proquest.com/lib/lshtmuk/detail.action?docID=214321.

A First Course in Probability. Ross S. Pearson, 8th Edition, 2008.

An Introduction to Medical Statistics, Bland J. M. OUP, 3rd Edition, 2000.

Harrell, F.E. Regression Modeling Strategies. Springer. [Extensive coverage of practical strategies for modelling data].

Dobson, A.J and Barnett, A.G. (2008) An Introduction to Generalized Linear Models, Third Edition. Chapman & Hall.

Other resources

Module information, including timetables, directed reading, lecture notes, practical instructions for each session will be made available via the Virtual Learning Environment (Moodle).

Teaching for Disabilities and Learning Differences

- Lectures will be recorded using Panopto in line with the LSHTM's policy on Lecture Recording.
- The module manual will be made available in advance of the start of the module and will be produced in accessible format.
- Slides will be made available in advance of each lecture or seminar and produced in accessible format.
- All material will be made available through Moodle.

The module-specific site on Moodle provides students with access to lecture notes and copies of the slides used during the lecture prior to the lecture (in pdf format). All lectures are recorded and made available on Moodle as quickly as possible. All materials posted up on Moodle areas, including computer-based sessions, have been made accessible where possible.

The LSHTM Moodle has been made accessible to the widest possible audience, using a VLE that allows for up to 300% zoom, permits navigation via keyboard and use of speech recognition software, and that allows listening through a screen reader. All students have access to "SensusAccess" software which allows conversion of files into alternative formats.

For students who require learning or assessment adjustments and support this can be arranged through the Student Support Services – details and how to request support can be found on the LSHTM Disability Support pages.