BIO STATISTICS COLLABORATION OF AUSTRALIA

Study Guide

Regression Modelling for Biostatistics 1 (RM1)

Semester 2, 2022

Prepared by:

Prof Stephane Heritier
School of Public Health and Preventive Medicine
Monash University

A/Prof Timothy Schlub
Sydney School of Public Health
Faculty of Medicine and Health
University of Sydney

Contents

Regression Modelling for Biostatistics 1 (RM1)
Semester 2, 2022

Contact details

Instructor Semester 1 2022	Instructor Semester 2 2022
Timothy Schlub Sydney School of Public Health Faculty of Medicine and Health University of Sydney (02) 9351 5992 tim.schlub@sydney.edu.au	Stephane Heritier School of Public Health and Preventive Medicine, Monash University (03) 9903 0054 stephane.heritier@monash.edu

If you have any general BCA queries, please contact: Karolina Kulczynska-Le Breton or Emily Higginson at the BCA Coordinating Office on 02 9562 5076/54 or via email to bca@sydney.edu.au

Background

The aim of this unit is to lay the foundation of "regression models" to analyse data from randomised or observational studies. "Regression" is a general term for measuring associations between an outcome and multiple covariates at once, allowing for the adjustment of confounding and effect modification. These methods are vital skills for you to be an effective practitioner of biostatistics as they are commonly used in a lot of health research. A suite of common regression models will be taught across this unit (Regression Models 1 (RM1)) and in the subsequent Regression Models 2 (RM2) unit. The skills taught in this unit (and in RM2) will be used by students for the remainder of their BCA studies.

In RM1 we will be focussing on regressions where the outcome variables are either continuously distributed (linear regression models), or are binary (logistic regression models). RM2 will then expand the logistic regression concepts introduced in RM1, and also include other regression models such as survival models in the framework of *general* linear models.

Unit summary

The focus of this unit will be on developing, validating and interpreting multivariable linear and logistic models. Emphasis will be placed on interpretation of results and checking the model assumptions. We also aim to provide a balance between theory and practice: mathematical proofs are not emphasised but mathematical literacy is

promoted to both establish a solid grounding in the main concepts and to enable students to build on the basic material covered here when needed with independent study. This subject combined with RM2 provides the core prerequisite knowledge in statistical regression modelling.

Many courses on regression emphasise the technical aspects of fitting and testing models (i.e. the Stata or R code). In practice however, the hardest challenges facing a biostatistician relate to issues of how to construct and interpret appropriate models in such a way that you (the biostatistician) can help provide reasonable answers to empirical research questions. Through the class discussions and activities, we hope to reinforce the message that good applied statistical work requires a lot of nuanced judgment, and decisions that are not necessarily either right or wrong. After all, statistics is essentially the art of handling uncertainty!

Workload requirements

As this is a new unit for 2022, we do not have good historic data on the time students spend studying the material and completing assessments each week. However we expect that most students would spend approximately on average 8-12 hours per week on this unit, with some students spending less time and others spending more. The expected workload consists of watching videos, guided readings, discussion posts, collaborative activities, independent study, tutorial participation and completion of assessment tasks.

Prerequisites

Mathematical Foundations for Biostatistics (MFB) and Epidemiology (EPI)
OR
Mathematical Background for Biostatistics (MBB) and Probability and Distribution Theory (PDT) and Epidemiology (EPI)

Co-requisites (must be taken before or concurrently with LMR)

Principles of Statistical Inference (PSI)

Learning Outcomes

At the completion of this unit students should be able to:

1. Explain the motivations for different regression analyses and be able to select and apply a suitable modelling approach based on the research aim
2. Analyse data using normal linear models, and be able to assess model fit and diagnostics
3. Analyse data using logistic regression models for binary data, and be able to assess model fit and diagnostics
4. Accurately interpret and manipulate mathematical equations that relate to regression analysis
5. Effectively communicate the outcomes and justification of a regression analysis

Unit content

The unit is divided into 12 weeks, summarised in more detail below. Each week involves 2-3 learning activities that may be either interactive, independent or collaborative in nature. This includes activities such as:

1. Watching recorded videos describing concepts and methods (visual and auditory learning)
2. Reading module notes or readings describing concepts and methods (linguistic learning)
3. Independently and collaboratively completing exercises and comparing solutions (solitary and social learning)
4. Participating in live interactive tutorials and discussions (social, visual and auditory learning)
5. Participating in asynchronous discussion (social and linguistic learning)
6. Independently investigating concepts and methods and presenting your findings to the class or a small group (solitary and social learning)
7. Working through provided computer code (linguistic and solitary learning)

The diversity of learning activities are designed to engage the different ways that students learn.

Study materials for all Modules are accessed from the eLearning unit site. Assignments and supplementary material such as datasets will be available within each Assignment item. Please note that we are not able to post copies of copyright material (journal articles and book extracts)—for these you will have to rely on your home university's library.

Recommended approaches to study

Students should begin each week by reading through the weekly instructions and completing the learning activities as directed – which will be different every week. As some of the activities are collaborative, you will get the best learning experience by keeping pace with the weekly schedule. Working behind the weekly schedule will mean you cannot participate in many of the interactive activities. At the end of each week's material, a summary document will be released that describes the important learning objectives of the week and with solutions to exercises. The purpose of this document is not to replace the learning activities, but rather to use for reflective study or review. There are several live online sessions throughout the semester and you will be expected to prepare for these and participate in the discussion if you are able to join. Recordings will be made for those who cannot attend.

Method of communication with coordinator(s)

The best method to get in contact with me is by email (for administrative or personal enquiries), by discussion board (for academic/content related enquiries) or by asking a question in the live tutorials. If you wish to discuss something over the phone, or via Zoom, it is best to first arrange a mutually suitable time by email.

Questions about administrative aspects or course content can be emailed to the coordinator, and when doing so please use "[RM1]:" in the Subject line of your email to assist in keeping track of our email messages. Coordinator/s will be available to answer questions as much as possible. However, please note that instructors are not necessarily available every day of the week and you should expect that it may take a business day or two to respond to questions or an email.

Canvas announcements will be the primary form of communication from me to you. Please check your canvas announcements regularly or set them up to automatically email notify you of new announcements.

Module descriptions

Below is an outline of the study weeks, followed by a timetable and assessment description table

Each week is scheduled to begin on a Monday and conclude on the Sunday of the following week.

Week	Topic
1	Simple linear regression
2	Checking assumptions in linear regression
3	Binary covariates, outliers and influential observations
4	Multiple linear regression application
5	Multiple linear regression theory
6	Interaction and multicollinearity
7	Assumption violations
8	Linear regression model building
9	Logistic regression
10	Confounding and interaction in logistic regression
11	Checking assumptions in logistic regression
12	Logistic regression model building

Unit schedule

Semester 2, 2022 starts on Monday 25 July

Week	Week commencing	Topic	Assessment
1	25 July	Simple linear regression	
2	1 August	Checking assumptions in linear regression	
3	8 August	Binary covariates, outliers and influential observations	Assignment 1 released
4	15 August	Multiple linear regression application	
5	22 August	Multiple linear regression theory	
6	29 August	Interaction and multicollinearity	Assignment 1 due Monday 29 August Assignment 2 released
7	05 September	Assumptions violations	
8	12 September	Linear regression model building	
9	19 September	Logistic regression	
Mid semester break 26 – 30 September			
10	3 October	Confounding and interaction in logistic regression	Assignment 2 due
11	10 October	Checking assumptions in logistic regression	Assignment 3 released
12	17 October	Logistic regression model building	
13	24 October		
14	31 October		Assignment 3 due on 31 October

Assessment

Assessment 1 and 2 are written assignments worth 30% each. Assessment 3 is a written assignment worth 40%

All assessments will be posted on the eLearning site together with an online Announcement broadcasting their availability.

Assessment name	Assessment type	Coverage	Learning objectives	Weight
Assessment 1	Written report	Weeks 1 – 4	LO1, LO2, LO4 and LO5	30%
Assessment 2	Written report	Weeks 1 – 8	LO1, LO2, LO4 and LO5	30%
Assessment 3	Written report	Weeks 1 - 12	LO1, LO2, LO3, LO4 and LO5	40%

In general, you are required to submit work typed in Word or similar. We strongly recommend you become familiar with equation typesetting software such as Microsoft's Equation Editor for algebraic work. You should not submit handwritten work unless you have sought approval from the unit coordinator to do so first. See the BCA Assessment Guide for guidelines on acceptable standards for assessable work.

Students are encouraged to discuss relevant topics in the Discussion Board. However, you should not post questions relating directly to Assessment. These should be emailed to the Unit Coordinator in the first instance.

Explicit solutions to assessable exercises should not be posted for others to use. Each student's submitted work must be clearly their own, with anything derived from other students' discussion contributions clearly attributed to the source.

Submission of assessments and academic honesty policy

All assessment material should be submitted via the relevant Assessment module in Canvas unless otherwise advised. Turnitin plagiarism detection is applied to all submissions. For detailed information, please see the BCA Assessment Guide, which includes links to the Academic Honesty policies at member universities. Please familiarise yourself with the procedures and policies at your home university. You will need to indicate your compliance with the plagiarism guidelines and policy at your home university.

Late submission of assessments and extension procedure

The standard BCA policy for late penalties for submitted work is a 5% deduction from the maximum possible mark for each day the assessment is late, up to a maximum of 10 days (including weekends and public holidays). Extensions are possible, but these need to be applied for (by email) as early as possible. The Unit Coordinator can approve extensions up to three days; for extensions beyond three days, you must apply to your home university, using their standard procedures.

Learning resources

The main text book for this unit of study is “Regression Methods in Biostatistics: Linear, Logistics, Survival, and Repeated Measures Models”, Eric Vittinghoff, David Glidden, Stephen Shiboski, Charles McCulloch. Second Edition. This textbook is available freely online through you home university library.

Software requirements and assumed knowledge

For this subject you can use either the Stata statistical package or R. If using Stata, the notes assume the use of release 12 or later of Stata. Most of the commands we use should work fine in older versions (as long as they are not too old!), although there was an important change relevant to RM1 with the introduction of "factor variables" in Stata 12. Importantly, whichever version you are using, please ensure that you have performed the online update to the latest update of that version. (Use the command update query.)

The notes for this course show both R and Stata code whenever possible. The textbook for this unit shows Stata code only, and so relevant equivalent code for R is shown in the notes.

For help with R, please see Learning R in the Student Resources site.

If you have not yet organised access to these packages, you should do so as soon as possible. This is a practical course which requires regular use of the relevant software; delays in gaining access to these packages may impact your ability to complete the course. Information on how to download R and RStudio, and access Stata can be found in the BCA Textbook and Software Guide.

Feedback

Our feedback to you:

The types of feedback you can expect to receive in this unit are:

- Formal individual feedback on submitted assessments
- Responses to questions posted on Canvas
- Interactive discussions of topics during live online tutorials
- Written feedback on individual investigations
- Peer feedback during collaborative activities

Your feedback to us:

One of the formal ways students provide feedback on teaching and their learning experience is through the BCA student evaluation survey at the end of each semester. The feedback is anonymous and provides the BCA with evidence of aspects that students are satisfied with and areas for improvement.

Unit changes, including response to recent student evaluation

Semester 1 2022 is the first instance this unit has run, so the course material it still very new. Only minor changes have been implemented between semester 1 and semester 2. We would like to have the benefit of a full year of teaching before making major changes. We have designed RM1 based on feedback from other Master of Biostatistics core units and on an external review of the Master of Biostatistics Program.