



---

## **Study Guide**

### **Regression Modelling for Biostatistics 1 (RM1)**

Semester 2, 2025

---

Prepared by:

Dr Liz Ryan & Prof Rory Wolfe  
School of Public Health and Preventive Medicine  
Monash University

Prof Armando Texeira-Pinto  
Sydney School of Public Health  
Faculty of Medicine and Health  
University of Sydney

Copyright © Sydney School of Public Health, University of Sydney & School of Public Health and Preventive Medicine, Monash University

## Contents

|  |    |
|--|----|
| Contact details .....  | 2  |
| Background .....   | 2  |
| Unit summary .....   | 2  |
| Workload requirements .....  | 3  |
| Prerequisites .....  | 3  |
| Co-requisites (must be taken before or concurrently with RM1) .....  | 3  |
| Learning Outcomes .....  | 3  |
| Unit content .....   | 4  |
| Recommended approaches to study .....                                | 5  |
| Method of communication with coordinator(s) .....                    | 5  |
| Module descriptions .....  | 6  |
| Unit schedule .....  | 7  |
| Assessment .....   | 8  |
| Submission of assessments and academic honesty policy .....          | 8  |
| Late submission of assessments and extension procedure .....         | 9  |
| Learning resources .....   | 9  |
| Software requirements and assumed knowledge .....                    | 9  |
| Feedback .....   | 11 |
| Unit changes, including response to recent student evaluations ..... | 11 |

## Regression Modelling for Biostatistics 1 (RM1)

Semester 2, 2025

### Contact details

| Instructors Semester 1 2025  | Instructors Semester 2 2025   |
|--|---|
| Prof. Armando Teixeira-Pinto (main)<br>Dr Farzaneh Boroumand (associate)<br>Sydney School of Public Health,<br>Faculty of Medicine and Health,<br>University of Sydney<br><br><a href="mailto:armando.teixeira-pinto@sydney.edu.au">armando.teixeira-pinto@sydney.edu.au</a><br><a href="mailto:farzaneh.boroumand@sydney.edu.au">farzaneh.boroumand@sydney.edu.au</a> | Dr Liz Ryan<br>Prof. Rory Wolfe<br>School of Public Health and Preventive Medicine,<br>Faculty of Medicine, Nursing and Health Sciences,<br>Monash University<br><br><a href="mailto:liz.ryan@monash.edu">liz.ryan@monash.edu</a><br><a href="mailto:rory.wolfe@monash.edu">rory.wolfe@monash.edu</a> |

If you have any general BCA queries, please contact the BCA Coordinating Office on (02) 8627 5033 or via email to [bca@sydney.edu.au](mailto: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 relationships between an outcome and multiple covariates at once, allowing adjustment for confounding and examination of effect modification. These methods are vital skills for you to be an effective practitioner of biostatistics as they are commonly used in health research. A suite of important regression models will be taught across this unit (Regression Modelling 1 (RM1)) and in the subsequent Regression Modelling 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 focus on regression models for outcome variables that are either continuously distributed (linear regression models), or are binary (logistic regression models). RM2 will then expand the logistic regression concepts introduced in RM1 in the framework of *generalised* linear models and also include other important regression techniques such as survival 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 model assumptions. We 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 unit combined with RM2 provides the core prerequisite knowledge in statistical regression modelling for the BCA courses.

Many units 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**

We expect that most students would spend 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 RM1)**

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 health research data using linear regression models, and be able to assess model fit and diagnostics
3. Analyse data using logistic regression models for binary outcome variables, 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 textbook 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 (solitary learning)
7. Working through provided computer code (linguistic and solitary learning)

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

Study materials for all Modules are accessed from Canvas. Assignments and supplementary material such as datasets will be available within each Assignment item on Canvas. 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 describing the important learning objectives of the week is displayed and solutions to the exercises provided. The purpose of this summary is not to replace the learning activities, but rather to be used for reflective study or review. There are several live online sessions (tutorials and Q&As) 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 us 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, please email the coordinator(s) to arrange a mutually suitable time.

Questions about administrative aspects or course content can be emailed to the coordinator(s) for the semester, 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.

**Canvas announcements** will be the primary form of communication from the RM1 coordinators 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 following Sunday.

| Week | Topic /Module   |
|------|---|
| 1    | Simple linear regression  |
| 2    | Checking assumptions in linear regression                                     |
| 3    | Binary covariates in linear regression, 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 introduction  |
| 10   | Confounding and interaction in logistic regression                            |
| 11   | Checking assumptions in logistic regression                                   |
| 12   | Logistic regression model building  |

## Unit schedule

Semester 2, 2025 starts on **Monday 28 July**

| Week   | Week commencing | Topic   | Tutorial/Q&A                     | Assessment  |
|--|-----------------|---|----------------------------------|---|
| 1  | 28 July         | Simple linear regression  | Tutorial (4pm AEST Tues 29 July) |   |
| 2  | 04 August       | Checking assumptions in linear regression                                     | Tutorial                         |   |
| 3  | 11 August       | Binary covariates in linear regression, outliers and influential observations | Tutorial                         | Assignment 1 released on Monday 11 August   |
| 4  | 18 August       | Multiple linear regression application  |                                  |   |
| 5  | 25 August       | Multiple linear regression theory   | Tutorial                         |   |
| 6  | 1 September     | Interaction and multicollinearity   |                                  | Assignment 1 due on Monday 1 September.<br><br>Assignment 2 released Friday 5 September |
| 7  | 8 September     | Assumptions violations  | Tutorial                         |   |
| 8  | 15 September    | Linear regression model building  | Q&A                              |   |
| 9  | 22 September    | Logistic regression introduction  | Tutorial                         | Assignment 2 due Friday 26 September  |
| <b>Mid semester break 29 September – 3 October</b> |                 |   |                                  |   |
| 10   | 6 October       | Confounding and interaction in logistic regression                            | Tutorial                         | Assignment 3 released Friday 10 October   |
| 11   | 13 October      | Checking assumptions in logistic regression                                   |                                  |   |
| 12   | 20 October      | Logistic regression model building  | Tutorial                         |   |
| 13   | 27 October      | Revision and assignment   | Q&A                              | Assignment 3 due on Friday 31 October   |
| 14   | 3 November      | Consolidation   |                                  |   |

The first tutorial will occur 4-5pm AEST Tuesday 29<sup>th</sup> July. The exact day and time of subsequent tutorials and Q&A sessions will be communicated after discussion with students in week 1.



## Assessment

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

All assessments will be posted on Canvas 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.

## Use of ChatGPT and other generative AI tools in assessment tasks

The assessment tasks in this Unit have been designed to be challenging, authentic and complex. Although individual assessment components may provide specific guidance regarding the use of generative AI tools (e.g., ChatGPT), successful completion of these components will require students to critically engage in specific contexts and tasks for which artificial intelligence will provide only limited support and guidance. In all cases, a failure to reference the use of generative AI may constitute student misconduct under the Student Code of Conduct of your University of enrolment. To successfully complete assessment tasks, students will be required to demonstrate detailed comprehension of their written submission independent of AI tools.

## 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 is “Regression Methods in Biostatistics: Linear, Logistics, Survival, and Repeated Measures Models. Second Edition” by Eric Vittinghoff, David Glidden, Stephen Shiboski, and Charles McCulloch. This textbook is available freely online through you home university library.

## Software requirements and assumed knowledge

For this unit 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. Whichever version you are using, we recommend that you perform the free online latest update of that version (use the command `update all`).

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
- 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 evaluations**

Some minor changes have been implemented between semester 1 and semester 2 in the form of correction of typographical errors, and providing additional explanations. We continue to revise and update the materials as any issues are raised.