BIO STATISTICS COLLABORATION OF AUSTRALIA

Study Guide

Mathematical Foundations for Biostatistics (MFB)

Semester 1, 2023

Prepared by:
Andrew Forbes
School of Public Health and Preventive Medicine
Monash University
E-mail: Andrew.Forbes@monash.edu

Rhys Bowden
School of Public Health and Preventive Medicine
Monash University
E-mail: Rhys.Bowden@monash.edu

[Initial materials for this unit were prepared in conjunction with University of Adelaide in 2022]

MONASH University

THE UNIVERSITY of ADELAIDE

Contents

Mathematical Foundations for Biostatistics (MFB)

Semester 1, 2023

Contact details

Prof Andrew Forbes	**Dr Rhys Bowden**
School of Public Health and Preventive Medicine Monash University	School of Public Health and Preventive Medicine Monash University
Tel: (03) 9903 0580	
E-mail: Andrew.Forbes@monash.edu	E-mail: Rhys.Bowden@monash.edu

Andrew Forbes and Rhys Bowden, both from Monash University, will be jointly coordinating the unit this semester, with Andrew Forbes taking the administrative role of overall coordinator.

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

To obtain a sound understanding of the statistical methods used in the design and analysis of medical and health studies, it is essential to have a thorough knowledge of the theoretical basis for these techniques. This unit will focus on developing foundational mathematical skills and then applying them to the study of probability and statistical distributions.

Context within the program

This unit, together with Principles of Statistical Inference (PSI), will provide the core prerequisite mathematical statistics background required for the study of later subjects in biostatistics.

Prerequisites

Nil

Co-requisites

Nil

Unit summary

This unit covers the foundational mathematical methods and probability distribution concepts necessary for an in depth understanding of biostatistical methods. The unit commences with an introduction to mathematical expressions, followed by the fundamental calculus techniques of differentiation and integration, and essential elements of matrix algebra. The concepts and rules of probability are then introduced, followed by the application of the calculus methods covered earlier in the unit to calculate fundamental quantities of probability distributions, such as mean and variance. Random variables, their meaning and use in biostatistical applications is presented, together with the role of numerical simulation as a tool to demonstrate the properties of random variables.

In terms of the structure of the MFB unit, approximately the first half of the unit is developing mathematical skills, with the second half being their application to probability and probability distributions. The material is interspersed with exercises for you to attempt and hence gain a deeper understanding of the theory and methods covered. There will be videos posted on Canvas to summarise content as you work through each module of the unit.

The first half of the unit introduces, in three modules, mathematical concepts such as functions, logarithms, exponentiation, inequalities, limits, derivatives, integration and matrices. For details on the contents within each module see below. For this part of the unit, you will only need to go through the provided module notes. There is no prescribed textbook for this component. Each module will have a section motivating it and describing the relevance of a particular mathematical concept and its connection to statistics. Worked examples are provided in each section. Practice exercises relating to each concept are also provided within each section. Solutions to practice exercises will be provided at the start of semester. Video lectures will be provided each week. These video lectures will introduce the relevant topics in the Module and highlight some of the concepts. In addition to these, two videos will introduce you to R programming and WolframAlpha. Notes detailing the use of R programming for integration and matrix manipulations is also provided at the start of weeks 3 and 6.

The second half of the unit makes extensive use of the ***prescribed*** textbook, “Mathematical Statistics with Applications“, by Wackerly, Mendenhall and Scheaffer (“WMS”, see below for details of this book) and you will be directed to readings from the WMS book and to complete selected exercises. Each reading from WMS will have a section in the notes motivating it and describing its importance, and therefore we anticipate that you will have our module notes and the WMS book side by side.

Worked solutions will be made available for the exercises during semester. A Student Solutions Manual for WMS exists and this provides very brief worked solutions for all odd numbered exercises in the WMS textbook; you may consider purchasing this text to enable you to have solutions to extra exercises that we don't set in MFB. This would allow you to undertake extra practice and in the past some students have said that this was helpful to them. Note that the solutions we provide in MFB for the exercises we assign from WMS are substantially more detailed than the solutions provided in Student Solutions Manual.

In MFB we make use of R, Stata and WolframAlpha (a computational intelligence software, details below) and we do not assume that you have previously used any of these. You are free to choose whether to use Stata or R for solving numerical exercises, or indeed a combination.

Workload requirements

The expected workload for this unit is 10-12 hours per week on average, consisting of guided readings, discussion posts, independent study, and completion of assessment tasks.

Learning Outcomes

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

1. Manipulate general mathematical expressions and inequalities
2. Understand and apply the essential elements of calculus, including differentiation and integration
3. Manipulate and evaluate matrix expressions and calculate inverses of matrices
4. Demonstrate an understanding of the meaning and laws of probability
5. Recognise common probability distributions and their properties
6. Apply calculus-based tools to derive key features of a probability distribution and properties of random variables, such as mean and variance
7. Demonstrate skills in simulation of random variables to illustrate and explain statistical concepts.

Unit content

The unit is divided into 7 modules, summarised in more detail below. Modules are of varying lengths, from one to four weeks. Each module will involve one or two weeks of study and includes the following material:

1. Module notes describing concepts and methods, and including some exercises of a more "theoretical" nature.

2. Selected readings and exercises from textbooks or other resources.

3. One or more online videos to summarise the module content

4. An online 'live' tutorial, with date and time chosen by a Doodle poll of student preferences. The recording of the tutorial will be made available online for students unable to make it to the live session, or for later viewing.

Study materials for all Modules are accessed from the Canvas MFB unit site. Assignments, and supplementary material such as datasets, will be made available on the Unit site 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 work through each module systematically, following the module notes and any readings referred to, and working through the accompanying exercises. *You will learn a lot more efficiently if you tackle the exercises systematically as you work through the notes.* You are encouraged to post any content-related questions to the Canvas site, whether they relate directly to a given exercise, or are a request for clarification or further explanation of an area in the notes. You should also work through all the computational examples in the notes for yourself on your own computer.

Solutions to the exercises in each module (except those to be submitted for assessment, as described below) are included in the study materials.

Method of communication with coordinators

Questions about administrative aspects or course content can be emailed to the coordinator, and when doing so please use "MFB:" in the Subject line of your email to assist in keeping track of our email messages. Coordinator/s will be available to answer questions related to the module notes and practical exercises, and to address any other issues that require clarification. However, please note that instructors are not necessarily available every day of the week and you should expect that it may take a day or so to respond to questions (possibly longer over weekends and during breaks!).

We strongly recommend that you post content-related questions to the Discussion Board in the unit site.

In 2023 we are using Canvas, the Learning Management system hosted by the University of Sydney. You will receive specific instructions on using the site this semester from the BCA Coordinating Office. There is also a "Getting Started" document available in the Student Resources site.

Module descriptions

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

Each module is scheduled to begin on a Monday and to conclude on a Sunday. **The due date for submission of the required exercises from each module is 11:59pm on the day indicated in the assessment table below.**

Module 1: **Numbers and functions (1 week)**

- Numbers and expressions
- Ratios and Percentages
- Powers, exponentials and logarithms
- Summations and Means
- Factorials
- Absolute values and inequalities
- Graphing Functions, use of Wolfram Alpha
- Families of functions
- The inverse of a function
- Solving quadratic equations

Module 2: Calculus (4 weeks)

- Limits
- Continuity of functions
- Tangent lines and rates of change
- Derivatives and techniques of differentiation
- The Product, Quotient and Chain Rules
- Derivatives of logarithmic and exponential Functions
- Analysis of Functions I: Increasing, Decreasing and Concavity
- Analysis of Functions II: Relative Extrema; Graphing Polynomials
- Introduction to Integration
- The definition of Area as a limit of rectangles
- Definite and Indefinite integrals
- Improper Integrals
- Systematic techniques for integration
- The gamma function
- Double Integrals
- Functions of Two or More Variables
- Partial Derivatives
- Newton's Method
- Taylor Series/Polynomials

Module 3: Matrices and determinants (1 week)

- Matrices and matrix operations
- Transpose and trace of a matrix
- Diagonal, Triangular and Symmetric Matrices
- Determinant of a matrix
- Inverse of a matrix, and properties
- Introduction to systems of linear equations

Module 4: Probability concepts (1 week)

- Use of set notation (Venn diagrams) including null set, union, subset, intersection, complement, and mutually exclusive.
- Definitions of events, , sample space, discrete sample space, and probability of an event.
- Calculation of the probability of an event by using combinatorial methods
- Application of conditional probability, independent events, the multiplicative and additive laws of probability, Bayes rule, the law of total probability.

Module 5: Discrete random variables (2 weeks)

- Definition of a random variable
- Probability distribution for a discrete random variable
- The Bernoulli, Binomial, Poisson and other distributions
- The expectation (mean) and variance of a discrete probability distribution
- Stata and R to perform probability calculations and graphs of distributions

Module 6: Continuous random variables (2 weeks)

- The cumulative distribution function and probability density function of a continuous random variable
- The expectation (mean) and variance of a continuous random variable given its density function
- Probability calculations using the Uniform, Normal and Lognormal distributions
- The expectation and variance of a transformation of a random variable
- The definition and use of moment generating functions for continuous random variables
- Simulating values from specified probability distributions

Module 7: Multiple random variables (2 weeks)

- Definitions of correlation, covariance and independence
- Joint, conditional and marginal distributions
- The Bivariate Normal distribution
- Expected value and variance of a linear function of random variables
- Marginal means and variances of random variables by iterated expectation
- The distribution of sums of independent random variables using the method of moment generating functions

Unit schedule

Semester 1, 2023 starts on Monday February 27. The mid-semester break is April 7-14, coinciding with the week of Easter.

Week	Week commencing	Module	Topic	Assessment
1	February 27	1	Numbers and functions	
2	March 6	2	Calculus (1)	
3	March 13	2	Calculus (2)	
4	March 20	2	Calculus (3)	**Assignment 1 due March 26 (15%)**
5	March 27	2	Calculus (4)	
6	April 3	3	Matrices	
7	April 10		Mid-semester break (April 7-14)	**Assignment 2 due April 16 (35%)**
	April 17	4	Probability	
8	April 24	5	Discrete random variables	
9	May 1			
10	May 8	6	Continuous random variables	
11	May 15	6		Assignment 3 due May 14 (15%)
12	May 22	7	Multiple random variables	
13	May 29	7		
14	June 5		***Work on assignment 4***	**Assignment 4 due June 11 (35%)**

Assessment

Assessment will include 4 written assignments worth 15%, 35%, 15%, 35% respectively, and will be released approximately 2 weeks prior to the due date. These assignments will be posted on the Canvas site together with an online Announcement broadcasting their availability.

Assessment name	Assessment type	Coverage	Learning objectives	Weight
Assignment 1	Assignment	Modules 1-2 (Concepts 1 & 2)	1,3	15%
Assignment 2	Assignment	Module 2 (Concepts 3 & 4)-Module 3	1,2,3,4	35%
Assignment 3	Assignment	Modules 4+5	1,4,5,6	15%
Assignment 4	Assignment	Modules 1-7	1,2,3,4,5,6,7	35%
Online quizzes	Non-assessed	Various modules	various	Not assessed

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 may submit neatly handwritten work, however please note that marks will potentially be lost if the solution cannot be understood by the markers due to unclear or illegible writing. Handwritten work should be scanned and clearly legible and collated into a ***single*** pdf file and submitted via the eLearning site. Submission of a set of mobile phone photos as separate files is not acceptable. 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, please avoid posting questions relating directly to assessable material. 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 must be clearly attributed to the source.

Submission 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.

A special note regarding "contract cheating" sites: Unfortunately there have been instances in the past of students using such websites to post assignment questions and receive solutions (usually for a fee). We have arrangements with these sites to identify the students posting questions or accessing the solutions, and such students will be referred to and face disciplinary processes at their home university.

Late submission and extension procedure
The policy for late penalties for submitted work is a 5% deduction from the earned 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

As stated earlier, the first half of the unit covers mathematical concepts and there is no prescribed textbook for this component. However, for the second half of the unit concerning probability and distributions, there is a prescribed textbook (described below) that will also be useful for later biostatistics units.

For the mathematical concepts component, we have created lecture notes that will be self-sufficient. There is also plenty of material available online concerning calculus and matrix algebra. The following web sites have proven useful to students in the past:

- Pauls' online Math notes- Free notes, examples, and practice problems for Algebra, Calculus 1,2, and 3 https://tutorial.math.lamar.edu/Classes/CalcI/CalcI.aspx
- Calculus.org- free practice problems, solutions and tips for exam preparation. Also provides addition resources. http://www.calculus.org/
- Khan Academy- Free short videos teach various concepts in a variety of disciplines https://www.khanacademy.org/math/differential-calculus
- Khan Academy- Free short videos on matrices. https://www.khanacademy.org/math/algebra-home/alg-matrices
- Matrix algebra tutorial https://stattrek.com/tutorials/matrix-algebra-tutorial.aspx

For the probability and distributions component of the unit, the prescribed textbook is:

Wackerley DD, Mendenhall W, Schaeffer RL. Mathematical Statistics with Applications. 7th edition. 2008 Thomson Learning, Inc. (Duxbury, Thomson Brooks/Cole) ISBN-13: 978-0-495-11081-1

This textbook is central to this component of the unit and you must have unrestricted access to this book. We refer to this textbook as "WMS" in the MFB material.

- There are several international editions of WMS available, however they are not identical and we do not recommend purchase of any edition which has a different ISBN to that of the edition in the BCA textbook and software guide, i.e., please only purchase ISBN-13 978-0-495-11081-1.
- Please be very careful if ordering this textbook online to ensure that the correct ISBN appears at each step of your ordering process – some websites advertise the above book and ISBN but their purchasing process, after displaying the correct book initially, subsequently skips to a different ISBN before you have completed the purchase.

The WMS textbook will also be useful for later biostatistics subjects because it also covers areas such as estimation, significance tests, and correlation and regression.

Other books which cover similar material for MFB and that we recommend are:

- Rosner B. Fundamentals of Biostatistics 4th edition.
  - *A textbook suitable for introductory courses in medical statistics that also touches on more advanced topics.*
- Larsen RJ & Marx ML. An Introduction to Mathematical Statistics and its Applications, Fourth Edition. 2006 Pearson International Edition.
  - *A direct competitor to WMS, this book is a useful source as an alternative to WMS; in general we prefer the WMS presentation and progression of topics but there are places where Larsen & Marx may be better.*

Software requirements and assumed knowledge

For this unit you will need to have access to the R or Stata software packages. In addition, you may also want to use WolframAlpha for learning the mathematics component and for checking calculus derivations. WolframAlpha runs in a web page so there is no software to obtain or download. No prior experience with any of R, Stata or WolframAlpha is required.

For the mathematical component in the first half of the unit you can choose between R and Wolfram alpha. For the probability and distributions component, which contains numerical calculations, you can choose between R and Stata, or use a mxiture. It does not matter which package you choose, and you can submit assessed work using either package.

For Stata, we expect most of you would be using versions Stata 15, 16 or 17. Although we are not aware of any major differences between Stata versions that affect the material, any minor issues will be pointed out in Canvas postings. Whichever version you are using, we recommend that you perform the online update to the latest update of that version. (Use the command `update query`). For R, we assume you are using at least R version 3.6, and we expect many of you will be using version 4.0.0 or later. You can check the version by typing `R.version`. We do not expect any differences in results between versions of R.

For help with R, please see Learning R in the Student Resources site.
For help with Stata, we can suggest introductory videos from Stata Corporation at https://www.stata.com/links/video-tutorials/ and we can suggest the "Tour of the Stata 17

interface" as a starting point (or version 16, 15 etc). There are also numerous introductory videos on YouTube obtained by typing "Introduction to Stata". You may also wish to sign up for the free webinars offered at https://surveydesign.com.au/webinars.html or look through the tips at https://www.techtips.surveydesign.com.au/blog/categories/stata.
For help with Wolfram alpha, we can suggest introductory video on YouTube https://www.youtube.com/watch?v=iYUGqX-qO_8. For more example on graphs, calculus, and elementary mathematics you can see the following webpage https://www.wolframalpha.com/.

Required mathematical background

Because this unit is an introductory foundational unit, the only background that is required is high school mathematics, and in particular the concepts of functions, differentiation and integration. Material will be presented starting at an elementary level, but the pace is rapid and some brushing up from previous studies will be advantageous.

Feedback

Our feedback to you:

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

- Formal individual feedback on submitted exercises assignments
- Feedback from non-assessed online quizzes
- Responses to questions posted in Canvas discussions

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

Minor errors in notes have been corrected. More signposting of the connection between calculus and probability sections has been added. An extra week has been added to Module 7.

Acknowledgments

Components of this unit are drawn from previous units entitled Mathematical Background for Biostatistics, and Probability and Distribution Theory. Modules 1-3 were initially developed for this unit by the University of Adelaide. We thank past developers and coordinators for the use of selected materials from these units.