BIO STATISTICS COLLABORATION OF AUSTRALIA

Study Guide

Principles of Statistical Inference (PSI)

Semester 2, 2023

Prepared by:

Dr. Erin Cvejic
Sydney School of Public Health
University of Sydney

THE UNIVERSITY OF SYDNEY

Contents

Contact details 2
Background 2
Context within the program 3
Prerequisites 3
Unit summary 3
Workload requirements 3
Learning Outcomes 3
Unit content 4
Recommended approaches to study 4
Method of communication with coordinator(s) 5
Module descriptions 6
Unit schedule 8
Assessment 9
Submission and academic honesty policy 9
A special note regarding “contract cheating” sites 10
Use of ChatGPT and other generative AI tools in assessment tasks 10
Late submission and extension procedure 10
Learning resources 10
Software requirements and assumed knowledge 11
Required mathematical background 12
Feedback 12
Unit changes, including response to recent student evaluation 12
Acknowledgments 13

Principles of Statistical Inference (PSI)
Semester 2, 2023

Contact details

Instructor

Dr Erin Cvejic
Sydney School of Public Health
University of Sydney
Ph: (02) 9351 5305 (however, email preferred due to availability)

Email: erin.cvejic@sydney.edu.au

Erin is a Senior Lecturer in Biostatistics, Program Director (Biostatistics), and Director of Public Health Methods at the University of Sydney School of Public Health. He is responsible for both the content and administration of the unit. One or more other academic biostatisticians from the Sydney School of Public Health may be assisting throughout the semester with involvement in online discussions and assessment marking.

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 email bca@sydney.edu.au

Background

A sound understanding of the basic principles of statistical inference, including the theory of statistical estimation and hypothesis testing, is necessary for students to gain a deeper understanding of methods used in the design and analysis of biomedical and epidemiological studies. Specifically, it verses students in the language of uncertainty. An understanding of the theoretical bases and drawbacks of common biostatistical techniques is essential for practising biostatisticians to be able to assess the validity of these techniques for particular studies, and to be able to modify those techniques where appropriate. In this unit of study (unit) students will develop a strong mathematical and conceptual foundation in the methods of statistical inference, which underlie many of the methods utilised in subsequent units of study, and in biostatistical practice.

Context within the program

This unit will develop a strong mathematical and conceptual foundation in the methods of statistical inference, with a particular focus on likelihood-based methods. These concepts and principles will underlie many of the methods utilised in subsequent units of study, as well as biostatistical practice in general.

Prerequisites
Mathematics Foundations for Biostatistics (MFB);
OR
Mathematical Background for Biostatistics (MBB) **AND** Probability and Distribution Theory (PDT)

Unit summary

The unit provides an overview of the concepts and properties of estimators of statistical model parameters, then proceeds to a general study of the likelihood function from first principles. This will serve as the basis for likelihood-based methodology, including maximum likelihood estimation, and the likelihood ratio, Wald, and score tests. Core statistical inference concepts including estimators and their ideal properties, hypothesis testing, p-values, confidence intervals, and power under a frequentist framework will be examined with an emphasis on both their mathematical derivation, and their interpretation and communication in a health and medical research setting. Other methods for estimation and hypothesis testing, including a brief introduction to the Bayesian approach to inference, exact and non-parametric methods, and simulation-based approaches will also be explored.

Workload requirements

The expected workload for this unit is **10-12 hours per week on average**, consisting of textbook and journal article readings, discussion board posts, independent study (lecture recordings and exercises), tutorial sessions, and completion of assessment tasks.

Learning Outcomes

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

1. Calculate and interpret important properties of point and interval estimators
2. Calculate and interpret p-values, power and confidence intervals correctly
3. Write a likelihood function
4. Derive and calculate the maximum likelihood estimate
5. Derive and calculate the expected information
6. Derive a Wald test, Score test, and likelihood ratio test

7. Use a Bayesian approach to derive a poster distribution
8. Calculate and interpret posterior probabilities and credible intervals
9. Apply and explain an exact method, non-parametric and sampling-based method

Unit content

The unit is divided into 6 modules, summarised in more detail below. Each module will involve approximately 2 weeks of study and includes the following material:

1. Recorded lectures on the theory content
2. One or two chapters from the textbook, which includes statistical theory and extended examples illustrating the statistical theory covered
3. A recorded lecture covering the extended example
4. Practical exercises, one of which is required to be submitted for assessment
5. Worked written and video solutions to non-assessed practical exercises
6. Discussion boards which can be used to ask questions and discuss module content and non-assessed exercises
7. Tutorial sessions for each module

With the exception of the textbook, study materials for all modules are downloadable from the eLearning (Canvas) unit site. Assignments and supplementary material, such as analysis datasets, will be posted to 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 begin each module by watching the lecture recording, reading through the relevant chapter of the textbook, and working through the extended example in parallel with the exercises. You are encouraged to post any content-related questions to the Canvas Discussion Boards, 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 any 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) will be posted online after the midway point of the allocated time period for the module. This is intended to encourage you to attempt the exercises independently before being given access to solutions.

Some of the exercises require computer simulations, and for these Stata and R code will be provided on Canvas. You are welcome to use any software you have available and are familiar with for the exercises (e.g., SAS, Matlab, Python), however code will not be provided for these packages and assistance will not be available if you get stuck.

Some exercises require the creation of graphs – it is strongly recommended that these are produced using statistical software, however a spreadsheet package (e.g., Excel) can be used. Irrespective of your choice of software used to create them, they must comply with the guidelines for reporting of statistical results found on the BCA student resources Canvas site.

Although a nominal period of 14 days is allocated to work on each module, students can ask questions about the material in any previous modules at any time during the semester.

Make the most of this unit by engaging with coordinators and your fellow students on the Discussion Board and in Tutorials. These are safe spaces to discuss the course material and related ideas and students are encouraged to make the most of them by engaging in respectful discussion.

Questions about Assignments should be directed to the coordinator in the first instance to avoid any Academic Honestly issues.

Method of communication with coordinator(s)

The Canvas site is the primary forum for communication between coordinators and students. It will also be used for posting all course material. The timetable below shows the dates when assignments will be made available and when all assessments are due. Please check the Canvas site regularly for new material and to keep up-to-date with class discussions.

Please post content-related questions to the relevant Discussion forum in the PSI Canvas site. You should be familiar with Canvas from previous BCA units. There is also useful information available on the BCA student resources Canvas site.

Questions about administrative aspects or course content can be emailed to the coordinator. Please use "PSI" 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.

Please note that instructors are not necessarily available every day of the week and you should expect that it may take up to **two business days** to respond to questions (possibly longer over weekends, during breaks, and NSW public holidays).

For matters of a personal nature, please contact the unit of study coordinator directly by email. Email is recommended to make initial contact and then a teleconference via Zoom can be arranged at a convenient time if required.

Module descriptions

Below is an outline of the study modules, followed by a timetable and assessment description table. Each module of this unit corresponds to a chapter in unit textbook. Each module is scheduled to begin on a Monday and conclude on the Sunday of the following week. **The due date for submission of the required exercises from each module is 11:59PM (Sydney Time) on the due date indicated below.**

Module 1: Estimation concepts (Chapter 2)

- Probability models and parameters
- Estimates and estimators
- Point and interval estimation
- Bias and efficiency
- Consistency and asymptotic efficiency

Module 2: Hypothesis testing concepts (Chapter 5)

- Null and alternative hypotheses
- Test statistics
- P-values
- Type I and Type II errors, significance level, and power
- Statistical significance and practical importance

Module 3: Likelihood and Estimation Methods (Chapters 3 and 4)

- Likelihood function
- Sufficiency
- Nuisance parameters
- Maximum likelihood estimation
- Statistical information
- Properties of maximum likelihood estimation

Module 4: Hypothesis testing methods (Chapter 6)

- Likelihood ratio tests
- Score tests
- Wald tests
- Relationship between the three tests
- Interval estimation based on the three tests

Module 5: Bayesian methods (Chapter 7)

- Basic concepts: subjective probability
- Bayes' rule, prior and posterior distributions
- Conjugate and non-informative prior distributions
- Analysis of simple binomial and normal models

Module 6: Further inference methods (Chapter 8)

- Exact methods
- Non-parametric methods
- Bootstrapping and other resampling methods

Unit schedule

Semester 2, 2023 starts on Monday 24th July (regardless of start date at home university).

Week	Week commencing	Module	Topic	Assessment
1	24th July	Module 1	Estimation concepts	M1 exercise due Sunday 6th August
2	31st July			
3	7th August	Module 2	Hypothesis testing concepts	M2 exercise due Sunday 20th August
4	14th August			
5	21st August	Module 3	Likelihood and estimation methods	***Assignment 1 released Friday 1st September*** M3 exercise due Sunday 3rd September
6	28th August			
7	4th September	Module 4	Hypothesis testing methods	M4 exercise due Sunday 17th September
8	11th September			
9	18th September	Module 5	Bayesian inference (Part 1)	**Assignment 1 due Tuesday 19th September**
	25th September	**Mid-semester Break**		
10	2nd October	Module 5	Bayesian inference (Part 2)	M5 exercise due Sunday 8th October
11	9th October	Module 6	Further inference topics	***Assignment 2 released Friday 20th October*** M6 exercise due Sunday 22nd October
12	16th October			
13	23rd October	**Revision and Assignment**		
	30th October			**Assignment 2 due 5th November**

Assessment

Assessment will include 2 written assignments worth 40% each, to be made available in the middle and at the end of the semester, and to be completed within approximately two weeks (including three weekends). These assignments will be posted on Canvas together with an online Announcement broadcasting their availability. In addition, students will be required to submit solutions to selected practical exercises (one from each module), worth a total of 20%, by deadlines specified throughout the semester (see table below).

Assessments are due by 11:59pm on the stated day.

Assessment name	Assessment type	Coverage	Learning objectives	Weight
Module 1 exercises	Assignment	Module 1	1,2	4%*
Module 2 exercises	Assignment	Module 2	1,2	4%*
Module 3 exercises	Assignment	Module 3	1,2,3,4,5	4%*
Assignment 1	Assignment	Modules 1-3	1,2,3,4,5	40%
Module 4 exercises	Assignment	Module 4	1,2,3,4,5,6	4%*
Module 5 exercises	Assignment	Module 5	1,2,3,4,7,8	4%*
Module 6 exercises	Assignment	Module 6	1,2,3,4,5,9	4%*
Assignment 2	Assignment	Modules 1-6	1,2,3,4,5,6,7,8,9	40%

* Your best five modules from six will each contribute 4% each towards the total 20% for the module exercises.

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. 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 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 student posting questions or accessing the solutions, and such students will be referred to and face disciplinary processes at their 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 artificial intelligence (AI) tools (e.g., ChatGPT), successful completion of these components will require students to critically engage in specific contexts and tasks for which AI 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 and extension procedure

The standard BCA 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

The textbook for this unit is:

Marschner, I.C.
Inference Principles for Biostatisticians
Chapman and Hall / CRC, 2014
ISBN 9781482222234 (hard cover)
ISBN 9780367576011 (paperback)
ISBN 9780429076244 (eBook)
http://www.crcpress.com/product/isbn/9781482222234

Note: there are a small number of minor typographical errors in the chapters used in this unit; a list of these will be provided on Canvas – please take note of these ***before*** reading the relevant chapters.

This book contains all the material that will be covered in this unit of study. Note, that you *may* have digital access to this text through you home university library – check this before you purchase a copy.

Other references books which you may find useful include:

1. Ross S. *A First Course in Probability*. MacMillan, 1988. **Background**
2. Azzalini A. *Statistical Inference: Based on the Likelihood*. Chapman and Hall, 1996. **Modules 1 – 4**.
3. Clayton D and Hills M. *Statistical Models in Epidemiology*. Oxford University Press, 1993. **Modules 1 – 4.**
4. Casella G and Berger RL. *Statistical Inference*. Wadsworth and Brooks/Cole, 1990. **Modules 1 – 4**.
5. McElreath R. *Statistical rethinking: A Bayesian Course with Examples in R and Stan.* Chapman and Hall / CRC, 2020. **Module 5**
6. Mood, A.M., Graybill, F.A. & Boes, D.C. (1963). *Introduction to the theory of statistics (3rd ed.).* McGraw-Hill. **Modules 1 – 4.**
7. Wackerley, D., Mendenhall, W., & Schaeffer RL. (2008). *Mathematical Statistics with Applications*. Wadsworth Group. **Modules 1 – 4.**

Additional readings (e.g., journal articles) may be recommended throughout the unit. Bibliographic details will be provided on Canvas.

Software requirements and assumed knowledge

For this subject you will need to have access to either Stata or R. The purpose of this unit is not to teach statistical computing (as this is covered by DMC Data Management and Statistical Computing). However, there are some exercises that rely on the use of simulation to help understand the concepts being taught which can only be achieved by using software. Stata and R code will be provided on Canvas which can be modified as required for exercises and assessments.

If you have not used Stata or R previously, it is highly recommended that you attempt to familiarise yourself with it prior to the beginning of semester. If you have not yet organised access to these packages, you should do so as soon as possible. Information on how to download R and RStudio, and access Stata can be found in the BCA Textbook and Software Guide.

Some students do struggle with the software elements of this unit. Please do not be afraid to ask for help from other students and instructors on Discussion Boards. Try not to allow any difficulties with software to obscure the basis of the course, which is to understand the principles of statistical inference. However, it is also important that practising biostatisticians can work in various software packages, so it is worthwhile making the effort to become proficient in at least one package.

Required mathematical background

Students should be familiar with the mathematical background covered as part of MFB (or MBB), including basic factorisation, rules for handling exponents and natural logarithms, differentiation and partial differentiation, and basic matrix manipulations (i.e., inverse of a matrix).

Feedback

Our feedback to you:

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

- Formal individual feedback on submitted exercises and assignments
- Responses to questions posted on Discussion Boards
- Informal feedback through discussion in Q+A sessions

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

PSI is delivered in both Semester 1 and Semester 2 each year. From Semester 1 2022 there were major changes to some of the core units in the BCA program. PSI has added additional material in Module 1 – this material was previously covered in PDT and therefore may be revision for some students. Other PSI modules have also been rearranged.

Based on feedback from previous deliveries, we have introduced recorded video lectures to complement the textbook readings, recorded worked video solutions to the non-assessed module exercises to further reinforce concepts, and provide the opportunity for live consultation (either in the form of tutorial or Q&A sessions, depending on module content) via videoconferencing to increase engagement and interactivity with the teaching team.

Acknowledgments

PSI has evolved over many deliveries, with valuable contributions from numerous colleagues and coordinators. The following people have been important in developing and delivering this unit and we thank them for their input:

- Liz Barnes
- Katrina Blazek
- Justin Zeltzer
- A/Prof Patrick Kelly
- Prof Ian Marschner