BIO STATISTICS COLLABORATION OF AUSTRALIA

Study Guide

Principles of Statistical Inference (PSI)

Semester 2, 2018

Prepared by:
Justin Zeltzer and Associate Professor Patrick Kelly
Sydney School of Public Health
The University of Sydney

THE UNIVERSITY OF SYDNEY

Contents

Instructor contact details

For **all enquiries about this unit**, contact the unit coordinator:

Justin Zeltzer

Sydney School of Public Health

Edward Ford Building (A27), University of Sydney, NSW 2006

email: justin.zeltzer@sydney.edu.au

web: http://sydney.edu.au/medicine/public-health/

In case of extended absence of the unit coordinator, the deputy coordinator is:

Associate Professor Patrick Kelly

Sydney School of Public Health

Edward Ford Building (A27), University of Sydney, NSW 2006

email: p.kelly@sydney.edu.au

phone: 02 9351 4369

web: http://sydney.edu.au/medicine/public-health/

For **enquiries about the BCA** and about the various degrees towards which this unit contributes, contact the BCA Executive Officer:

Erica Jobling

NHMRC Clinical Trials Centre, University of Sydney, NSW 2006

email: bca@ctc.usyd.edu.au

phone: 02 9562 5076

fax: 02 9565 1863

web: http://www.bca.edu.au/

For **enquiries about your degree program**, contact the university through which you are enrolled.

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. This unit of study (unit) provides the core prerequisite knowledge in statistical inference, which will subsequently be built upon in other subjects.

Unit summary

This unit will introduce and review core concepts of statistical inference, including estimators, confidence intervals, type I & II errors and p-values. The emphasis will be on the practical interpretation of these concepts in biostatistical contexts, including an emphasis on the difference between statistical and practical/clinical significance. Concepts in classical estimation theory, including bias and efficiency will be discussed. The unit will provide a general study of the likelihood function, which will be used as a basis for the study of likelihood-based methodology, including maximum likelihood estimation and inference based on likelihood ratio, Wald and score test procedures. The Bayesian approach to statistical inference will be briefly studied and contrasted with the classical frequentist approach.

Prerequisite courses and assumed knowledge

Mathematical Background for Biostatistics (MBB), Probability and Distribution Theory (PDT)

PSI builds upon the material covered in Probability and Distribution Theory (PDT). You may find it useful to refer back to your PDT notes. The first two chapters and the appendix of the PSI textbook contain information that will be helpful for PSI – it is recommended that you read those chapters early in the semester and refer to the appendix as required.

Learning outcomes

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

1. Write a likelihood function
2. Derive and calculate the maximum likelihood estimate
3. Derive and calculate the expected information
4. Calculate and interpret p-values, power and CIs correctly
5. Derive a Wald test, Score test and likelihood ratio test
6. Use a Bayesian approach to derive a poster distribution
7. Calculate and interpret posterior probabilities and credible intervals
8. Apply and explain an exact method, non-parametric and sampling-based method.

Unit content

The unit is divided into 6 modules, summarized in more detail below. Each module will involve approximately 2 weeks of study and will usually include the following materials:

1. A chapter from the textbook, which includes statistical theory and an extended example illustrating the statistical theory covered.
2. A lecture on the theory and a lecture on the extended example
3. Practical exercises to be completed
4. A discussion board to ask lots of questions and post up solutions to non-assessed exercises.

Students should begin each module by reading through the relevant chapter of the text and work through the extended example in parallel with the exercises.

Lectures will be provided describing the material in each module and the extended examples where applicable. This year there have been substantial changes to PSI from previous deliveries of the unit. These include replacing module notes with a textbook, the introduction of new unit material and the re-ordering/numbering of the previous existing modules. We have recording new lectures, but we are also re-using some previously recorded lectures, so please be aware that for these lectures that references to the module number and/or page numbers may not be directly applicable.

Each module contains a number of exercises, one of which is required to be submitted for assessment per module. Other provided exercises are not assessable, but will assist greatly the student's learning. Discussion boards will be used by students to corroborate on answers for the non-assessed problem questions.

Module descriptions

Each module of this unit corresponds to a chapter in the Marschner textbook (see below for details).

Module 1: Likelihood
- Likelihood function
- Sufficiency
- Nuisance parameters
- Approximate likelihood

Module 2: Estimation methods
- Maximum likelihood estimation
- Fisher information
- Properties of maximum likelihood estimation

Module 3: Hypothesis testing concepts
- Null and alternative hypotheses
- Test statistics
- P-values
- Type I & II errors, significance level and power
- Statistical significance and practical significance
-

Module 4: Likelihood based methods
- Likelihood ratio tests
- Score tests
- Wald tests
- Relationship between the three tests
- Interval estimation based on the three tests

Module 5: Bayesian methods
- Basic concepts: subjective probability
- Bayes' rule, prior and posterior distributions
- Conjugate and non-informative prior distributions
- Analysis of simple binomial & normal models

Module 6: Further inference methods
- Exact methods
- Non-parametric methods
- Bootstrapping and other resampling methods

Unit schedule

The following timetable describes the timing of the study modules and assessment tasks for Principles of Statistical Inference. It is the intention that students will work through the material for each module, including completion of practice exercises by the due date of the module assessment task. **All assessment tasks are due by 11:59pm on the Tuesday stated date.**

Week	Topic	Tuesday date	Assessment
1	Module 1	31 July	
2		7 August	
3	Module 2	14 August	Module 1 exercise due
4		21 August	
5	Module 3	28 August	Module 2 exercise due
6		4 September	
7	Module 4	11 September	Module 3 exercise due Assignment 1 handed out
8		18 September	
	Mid-semester Break	25 September	Assignment 1 due
9	Module 5	2 October	Module 4 exercise due
10		9 October	
11	Module 6	16 October	Module 5 exercise due
12		23 October	
13		30 October	Module 6 exercise due Assignment 2 handed out
14		6 November	
15		13 November	Assignment 2 due

Assessments

There are two assignments worth 40% each. In addition, one designated exercise from each module will be required to be submitted as part of the unit assessment. These assessable exercises are worth 20% in total. Your best five assessed exercises, from six, will count towards the 20%, i.e. your best five modules assessments will be worth 4% each. Your lowest scoring module assessment will not count towards your final assessment.

The due dates for assessments are outlined below, as well as in the semester timetable.

Assessment	Due Date	% marks
Module 1 exercise	14 August	4*
Module 2 exercise	28 August	4*
Module 3 exercise	11 September	4*
Assignment 1	25 September	40
Module 4 exercise	2 October	4*
Module 5 exercise	16 October	4*
Module 6 exercise	30 October	4*
Assignment 2	13 November	40
Total		100

* Your best five modules from six will be worth 4% each.

Module exercises are located within the Canvas Module folder. Assignments will be posted online. All assessments are due on **Tuesday by 11:59pm**.

Assessed exercises and assignments should be submitted on the Canvas website. Please consult the **BCA Assessment Guide** for details about submitting your assignments, and guidelines for written work. For PSI submissions:

- Your assessment submission should be **one** WORD or PDF document. Multiple files/documents will **not** be accepted.
- Do **not** send EXCEL spreadsheets.
- Solutions for both exercises and assignments should include any formulae that you have used to arrive at your conclusion. Marks are awarded for correct formulae and working.
- If using WORD, please use the **equation editor** for writing mathematical formulae.
- Hand written solutions are acceptable.
- Marks will be deducted if incorrect notation is used.
- Some exercises require the creation of graphs – these can be done in statistical software or a spreadsheet and must comply with the guidelines for reporting of statistical results found on the BCA website:

http://www.bca.edu.au/currentstudents.html.

Submission of assessments and academic honesty policy

See the section on **Academic Honesty** of the **BCA Assessment Guide**. Please ensure that you familiar with the academic honesty policy associated with the university in which you are enrolled.

All material submitted for assessment in PSI must be entirely your own work**.**

You should submit all your assessment material via Canvas unless otherwise advised. The use of Turnitin for submitting assessment items has been instigated within unit sites. For more detail please see pages 3-5 the **BCA Assessment Guide**. When submitting assessments using Turnitin you will need to indicate your compliance with the plagiarism guidelines and policy at your university of enrolment before making the submission.

Acknowledgment of receipt of submitted exercises/assignments will be automatically generated in Canvas. It is the responsibility of the student to ensure that the assessment item is received by the unit coordinator by the due date.

Late submission of assessments and extension procedure

We adhere to standard BCA policy for late penalties for submitted work, i.e. a 5% deduction from the earned mark for each day the assessment is late, up to a maximum of 50%. Extensions are possible, but these need to be applied for (by email) as early as possible. The Unit Coordinator is not able to approve extensions beyond three days; for extensions beyond three days you need to apply to your home university, using their standard procedures.

Non-assessable Exercises

There will also be a range of non-assessable exercises provided each week. Discussion boards have been provided for each question so that students can submit their answers and collectively arrive at a set of answers for these questions. **Official answer guides to these questions will not be provided though hints and feedback may be given where necessary.** All students are implored to participate in these discussion threads. This has been shown in previous semesters to increase student interaction with the course content and prevents some students from waiting for answer guides before starting their study. PSI is a difficult course and has a considerable workload. If you fall behind it may be difficult to catch up. If you have difficulty keeping up with the workload for any reason, please contact the coordinator as soon as possible.

Textbook and other reference books

The textbook for this unit is:

Marschner I.C.
Inference Principles for Biostatisticians.
Chapman and Hall/CRC, 2014.
ISBN 9781482222234
http://www.crcpress.com/product/isbn/9781482222234
It contains all the material that will be covered in this unit.

Other reference books 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. Berger RL. *Statistical Inference.* Wadsworth and Brooks/Cole, 1990. **Modules 1 – 4.**
5. Gelman A, Carlin JB, Stern HS and Rubin DB. *Bayesian Data Analysis.* Chapman and Hall, 1995. **Module 5.**
6. Lee PM. *Bayesian Statistics.* Oxford University Press, 1989. **Module 5.**
7. Mood, A.M., Graybill, F.A. & Boes, D.C. (1963). *Introduction to the theory of statistics 3rd edition.* McGraw-Hill. **Modules 1 – 4.**
8. Wackerley, D., Mendenhall, W., & Schaeffer RL. (2008). *Mathematical Statistics with Applications.* Wadsworth Group. **Modules 1 – 4.**

Many statistical textbooks are not entirely devoted to inference, but have several sections on inference, which may not be as theoretical as the books above. Two of many are:

Altman DG. *Practical Statistics for Medical Research* Chapman and Hall, 1991
Fisher LD, van Belle G. *Biostatistics A Methodology for the Health Sciences* Wiley, 1993.

Method of delivery and communication

Canvas

We will use the Canvas PSI unit site as the main means of communication. A quick introduction to Canvas has been prepared by the University **here**, with more detailed help resources found on the **Student Canvas Guide**.

The Canvas PSI unit site should be the primary mode for asking questions, via the Discussion Board. We encourage you to use the Discussion Board as a forum for asking questions, posting comments and answering other students' questions. Interaction between students has been very successful in the past. Students are often able to resolve issues amongst themselves and pedagogic research suggests that this is often a better form of learning than relying solely on the instructors. The instructors will regularly monitor the discussions and contribute when needed.

Instructors will respond when required in a timely manner between Monday and Friday (within 48 hours from posting). Please note that instructors will not be available to reply to email or Discussion Board postings during the weekends.

The Canvas PSI unit site will also be used for posting additional course materials.

Email and phone

The unit coordinator will be available by e-mail and phone to answer any queries. Queries related to the module notes and practical exercises which are posted on **Canvas will be given priority over those sent via e-mail**.

To assist in keeping track of e-mails, please use "PSI" in the subject line of your email.

Software

The purpose of the course is not to teach statistical computing. However, there are several exercises that rely on the use of simulation to help understand the concepts being taught.

The recommended and supported software for this unit is either SAS or STATA. Whenever you will be required to use statistical software, the necessary code will be downloadable from the Canvas PSI unit site. The code can be run on your computer, and usually you will only need to change input values for exercises and assignments.

Some students sometimes struggle with the software. Please do not be afraid to ask for help from the other students and instructors on the Discussion Board. Try not to allow any difficulties with software to obscure the basis of the course, which is to understand the principles of statistical inference