BIO STATISTICS COLLABORATION OF AUSTRALIA

Study Guide

Principles of Statistical Inference (PSI)

Semester 1, 2018

Prepared by:
Liz Barnes
NHMRC Clinical Trials Centre
University of Sydney

THE UNIVERSITY OF SYDNEY

Contents

Instructor contact details

Liz Barnes
NHMRC Clinical Trials Centre
University of Sydney
Locked Bag 77
Camperdown NSW 1450
Ph: (02) 9562 5303
Fax: (02) 9565 1863 (Attn: Liz Barnes)
Email: Liz.Barnes@ctc.usyd.edu.au

Liz Barnes is a biostatistician at the NHMRC Clinical Trials Centre at the University of Sydney. She is responsible for both content and administration of this unit. One or more other biostatisticians from the CTC may also assist with marking of assessments and responding to eLearning queries throughout the semester.

Background

A sound understanding of the basic principles of statistical inference, including a grounding in 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. An understanding of the theoretical basis for 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 units.

Unit summary

The unit will introduce and review core concepts of statistical inference, including estimation, hypothesis testing, 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. 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. Further inference topics will also be introduced.

Workload requirements

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

Pre-requisites

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 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:

1. Have a deeper understanding of fundamental concepts in statistical inference and their practical interpretation and importance in biostatistical contexts
2. Understand the theoretical basis for frequentist and Bayesian approaches to statistical inference
3. Be able to apply likelihood-based methods of inference, with particular reference to problems of relevance in biostatistical contexts

Unit content

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.
A list of known errors in the textbook is available on the PSI eLearning site.

The unit is divided into 6 modules, summarised in more detail below. Each module will involve approximately 12-14 days of study and generally includes the following material:

1. A chapter from the Marschner textbook describing concepts and methods, and including some exercises of a more "theoretical" nature and (except Module 6/Chapter 8) an extended example that applies the theory in a practical context.
2. A number of exercises, one of which is required to be submitted for assessment.
3. Recorded lectures covering the material in each module and extended example

Apart from the textbook, study materials for all modules are downloadable from the eLearning unit site. Assignments and supplementary material, such as analysis datasets, will be posted to the unit site.

Recommended approaches to study

Students should work through each module systematically, following the textbook chapter and working through the accompanying exercises. You are encouraged to post any content-related questions to eLearning, whether they relate directly to a given exercise, or are a request for clarification or further explanation of an area in the textbook. You should also work through any computational examples in the textbook 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 at 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 SAS and Stata code will be provided on eLearning. You are welcome to use any other software you have available and are familiar with for the exercises, however code will not be provided and assistance may not be available. 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

Although a nominal period of 12-14 days is allocated to work on each module, students can ask questions about the material in any of the modules at any time during the semester.

Method of communication with coordinator(s)

The eLearning website is the primary forum for communication between co-ordinators and students. It will also be used for posting all course materials. The timetable below shows the dates when assignments will be made available. Please check the website regularly for new material and to keep up with class discussions.

Please post content-related questions to the relevant Discussion forum in the PSI eLearning site. You should be familiar with the eLearning system from previous BCA units, and will receive any specific instructions on using the eLearning site this semester from the BCA Coordinating Office. There is also a “Getting Started” document available on the Student Resources page of the BCA website.

Questions about administrative aspects or course content can be emailed to the coordinator, and when doing so 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. 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 NSW public holidays).

For personal matters please email or phone the co-ordinator.

Module descriptions

Each module of this unit corresponds to a chapter in the Marschner textbook.

Module 1 (Chapter 3) Likelihood

- Likelihood function
- Sufficiency
- Nuisance parameters
- Approximate likelihood

Module 2 (Chapter 4) Estimation methods

- Maximum likelihood estimation
- Statistical information
- Properties of maximum likelihood estimation

Module 3 (Chapter 5) Hypothesis testing concepts

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

Module 4 (Chapter 6) Hypothesis testing methods

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

Module 5 (Chapter 7) Bayesian methods

- 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 (Chapter 8) Further inference methods

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

Unit schedule

Semester 1, 2018 starts on Monday 26th February

Dates	Content	Assessment	Assessment due*
Feb 26 – Mar 11	Module 1	Module 1 exercise	Sunday Mar 11
Mar 12 – Mar 25	Module 2	Module 2 exercise	Sunday Mar 25
Mar 26 – Apr 8	Module 3	Module 3 exercise	Sunday Apr 8
Fri Apr 6 (5pm)		Assignment 1	Sunday Apr 22
Apr 23 – May 4	Module 4	Module 4 exercise	Friday May 4
May 5 – May 16	Module 5	Module 5 exercise	Wednesday May 16
May 17 – May 28	Module 6	Module 6 exercise	Monday May 28
Fri May 25 (5pm)		Assignment 2	Monday Jun 11

*Exercises and assignments must be submitted by 12 midnight (end of the day) on the due date.

Assessment

Assessment will include two 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 the eLearning site on the date listed above.
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).

Assessment name	Assessment type	Coverage	Weight
Module 1 submissible exercise	Assignment	Module 1	4%*
Module 2 submissible exercise	Assignment	Module 2	4%*
Module 3 submissible exercise	Assignment	Module 3	4%*
Assignment 1	Assignment	Modules 1-3	40%
Module 4 submissible exercise	Assignment	Module 4	4%*
Module 5 submissible exercise	Assignment	Module 5	4%*
Module 6 submissible exercise	Assignment	Module 6	4%*
Assignment 2	Assignment	Modules 1-6	40%

* **Your best five module marks will each contribute 4% towards the total of 20% for the module exercises**

In general you are required to submit your work typed in Word or similar (e.g. using Microsoft's Equation Editor for algebraic work) and we strongly recommend that you become familiar with equation typesetting software such as this. If extensive algebraic work is involved you may submit neatly handwritten work, however please note that marks may be lost if the solution cannot be understood by the markers due to unclear or illegible writing. This handwritten work should be scanned and collated into a single pdf file and submitted via the eLearning site. See the BCA Assessment Guide document for specific guidelines on acceptable standards for assessable work.

For all assessments please follow the guidelines for reporting of statistical results found at this website: http://www.bca.edu.au/currentstudents.html.

The instructors will generally avoid answering questions relating directly to the assessable material until after it has been submitted, but we encourage students to discuss the relevant parts of the chapter among themselves, via eLearning. However **explicit solutions to assessable exercises should not be posted for others to use**, and 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

You should submit all your assessment material via eLearning 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 Student Assessment Guide.
The BCA pays great attention to academic honesty procedures. Please be sure to familiarise yourself with these procedures and policies at your university of enrolment. Links to these are available in the BCA Student 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.

Late submission of assessments and extension procedure

While there is some flexibility in the submission dates for module exercises (late submissions are not usually penalised but their marking is likely to be delayed), **the deadlines for submission of assignments 1 and 2 are not negotiable**. For late assignment submissions 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.

Learning resources

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.

Software

The purpose of the course is not to teach statistical computing. However, there are some 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 PSI eLearning website. The code can be run on your computer, and usually you will only need to change input values for exercises and assignments. If you have not used your chosen software package before, it is highly recommended that you attempt to familiarise yourself with the package prior to the beginning of semester.

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. However, it is also important that practising biostatisticians are able to work in various software packages so it is worthwhile making the effort to become proficient in at least one package.

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 Blackboard

Your feedback to us:

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