BIO STATISTICS COLLABORATION OF AUSTRALIA

Study Guide

Principles of Statistical Inference (PSI)

Semester 1, 2021

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

THE UNIVERSITY OF SYDNEY

Contents

Instructor contact details

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

Liz Barnes

NHMRC Clinical Trials Centre

Medical Foundation Building

92-94 Parramatta Road, Camperdown NSW 2050

Email: Liz.Barnes@ctc.usyd.edu.au

Phone: +61 2 9562 5303

Liz is a biostatistician and research fellow at the NHMRC Clinical Trials Centre (CTC) at The University of Sydney. She is responsible for both the content and administration of this unit. One or more other biostatisticians from the CTC may be assisting throughout the semester with marking of assessments and responding to enquires made through eLearning.

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 be built upon in later 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 board posts, independent study and completion of assessment tasks.

Prerequisites

Mathematic Background for Biostatistics (MBB)
Probability and Distribution Theory (PDT)

PSI builds extensively 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 – we recommended that you read those chapters early in the semester and refer to the appendix as required throughout the unit.

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 confidence intervals 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

There are 6 modules, summarized in more detail below. Each module will involve 10-14 days 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 questions about the module material and non-assessed exercises.

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.

Recommended approaches to study

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. 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 notes. You should also work through all of 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) 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 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 (e.g., R, SAS), however code will not be provided for these packages and assistance may not be available. Some exercises require the creation of graphs – these can be done in statistical software or a spreadsheet package (e.g., Excel) and must comply with the guidelines for reporting of statistical results found on the BCA website: BCA guide for reporting statistical results

Although a nominal period of 10-14 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

The eLearning website is the primary forum for communication between co-ordinators and students. It will also be used for posting all course material. The timetable below shows the dates when assignments will be made available. Please check the website 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 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 content 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, during breaks, and NSW public holidays).

For personal matters, please email or phone the unit of study coordinator.

Module descriptions

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

Module 1: Likelihood (Chapter 3)
- Likelihood function
- Sufficiency
- Nuisance parameters
- Approximate likelihood

Module 2: Estimation methods (Chapter 4)
- Maximum likelihood estimation
- Statistical information
- Properties of maximum likelihood estimation

Module 3: Hypothesis testing concepts (Chapter 5)
- Null and alternative hypotheses
- Test statistics
- P-values
- Type I & II errors, significance level and power
- Statistical significance and practical importance
-

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 & normal models

Module 6: Further inference methods (Chapter 8)
- 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 stated date.

Semester 1, 2020 starts on Monday 2nd March

Dates	Content	Assessment	Assessment due*
Mar 01 – Mar 14	Module 1	Module 1 exercise	Sunday Mar 14
Mar 15 – Mar 28	Module 2	Module 2 exercise	Sunday Mar 28
Mar 29 – Apr 11	Module 3	Module 3 exercise	Sunday Apr 11
Fri Apr 09 (5pm)		Assignment 1	Sunday Apr 25
Apr 26 – May 07	Module 4	Module 4 exercise	Friday May 07
May 08 – May 19	Module 5	Module 5 exercise	Wednesday May 19
May 20 – May 31	Module 6	Module 6 exercise	Monday May 31
Fri May 28 (5pm)		Assignment 2	Monday Jun 14

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. These assignments will be posted on the eLearning site on the dates specified in the Unit Schedule. 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 above).

Assessment name	Assessment type	Coverage	Due Date	Weight
Module 1 exercise	Assignment	Module 1	Sunday 14 March	4%*
Module 2 exercise	Assignment	Module 2	Sunday 28 March	4%*
Module 3 exercise	Assignment	Module 3	Sunday 11 April	4%*
Assignment 1	Assignment	Modules 1-3	Sunday 25 April	40%
Module 4 exercise	Assignment	Module 4	Friday 07 May	4%*
Module 5 exercise	Assignment	Module 5	Wednesday 19 May	4%*
Module 6 exercise	Assignment	Module 6	Monday 31 May	4%*
Assignment 2	Assignment	Modules 4-6	Monday 14 June	40%
Total				100

*** Your best five modules from six 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 will potentially be lost if the solution cannot be understood by the markers due to unclear or illegible writing. Please scan and collate handwritten work into a single pdf file and submit via the eLearning site. See the BCA Assessment Guide document for specific guidelines on acceptable standards for assessable work. The instructors will generally avoid answering questions relating directly to the assessable material until after submission, but we encourage students to discuss the relevant parts of the notes 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

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 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 hard cover
ISBN 9780367576011 paperback
ISBN 9780429076244 eBook

This book contains all of 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.

There is a list of known errors in the textbook on the eLearning website.

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. Berger RL. Statistical Inference. Wadsworth and Brooks/Cole, 1990. Modules 1 – 4.
5. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A and Rubin DB. Bayesian Data Analysis (3 rd ed.). Chapman and Hall/ CRC Press, 2013. 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 ed.). 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

Software

The purpose of this unit 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 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 will only need to change input values for exercises / assignments. If you have not used Stata previously, it is highly recommended that you attempt to familiarise yourself with it prior to the beginning of semester. Some students do struggle with the software. 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 obscure the basis of the course, which 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.

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
- Responses to questions posted on Discussion boards in eLearning

Your feedback to us:

One of the formal ways that students can 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.