



Study Guide

Principles of Statistical Inference (PSI)

Semester 1, 2017

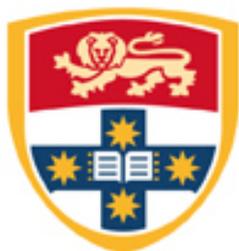
Prepared by:

Liz Barnes

NHMRC Clinical Trials Centre

University of Sydney

Copyright © NHMRC Clinical Trials Centre, University of Sydney



THE UNIVERSITY OF
SYDNEY

Co-ordinator:

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

Contents

Background	3
Unit Summary	3
Objectives	3
Assumed Knowledge	4
Unit Materials	4
Outline of Unit Modules	5
Method of Delivery and Communication	6
eLearning	6
Email and phone communication	6
Software	7
Textbook	7
Timetable and due dates	8
Assessment	9
Assignments	9
Exercises	9
Submission of assessable material	10
Essentials for submission	10
Contact details	11

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 subjects.

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.

Objectives

At the completion of this unit the student will:

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 ; and
3. Be able to apply likelihood-based methods of inference, with particular reference to problems of relevance in biostatistical contexts.

Assumed Knowledge

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.

Unit Materials

The unit is divided into 6 modules, summarized in more detail below. Each module corresponds to a chapter in the Marschner textbook. Lectures will be provided describing the material in each module and the extended examples where applicable.

There have recently been substantial changes to PSI from previous deliveries of the unit. These include replacing module notes with the textbook and re-numbering of modules. Semester 1 2016 will be the third delivery of the unit with the changes, and although the co-ordinators have tried hard to detect and fix all problems, there are likely to still be issues that need to be addressed. If you see anything in the notes or lectures that doesn't make sense please let us know, and please be patient while we continue to refine the materials.

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

Each module contains a number of exercises, one of which is required to be submitted for assessment.

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.

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>.

Outline of Unit Modules

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
- Statistical 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

Method of Delivery and Communication

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

eLearning

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.

The eLearning website should be the primary forum for asking questions. In later units students may be given marks for their contribution to discussions on eLearning. We will not use that practice here, but encourage you to use the Discussions section in the PSI area of eLearning as a forum for discussion of both content and the non-assessable exercises. This method of interaction between students has been successful in the past, to help students resolve issues amongst themselves. The coordinators will monitor the discussions and contribute when necessary.

Solutions to non-assessable exercises will be posted on eLearning, approximately half way through the time allocated for the module. It is strongly advised that you work through the exercises as much possible before accessing the solution.

Email and phone communication

The co-ordinator will be available by e-mail and phone to answer questions related to the module notes and practical exercises, or to address any other issues that require clarification, however, **queries on eLearning will be given priority over e-mail questions.** Questions about administrative matters should be directed to Liz Barnes via phone or email.

Every effort will be made to intervene in eLearning discussions, if necessary, in a timely manner between Monday and Friday (24 to 48 hours from posting, depending on the nature of the problem). When an answer/explanation cannot be provided within 24 hours, the coordinators will post an acknowledgment of the problem and respond as soon as possible afterwards. Please note that instructors will not be routinely available to reply to email or eLearning questions during weekends and NSW public holidays.

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, then you should 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.

Textbook

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.

Timetable and due dates

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 end date of the module. Due dates for assessment tasks are also shown.

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**. Late submissions may be penalised according to the guidelines under “Assignments” below.

Mar 6 – Mar 19 Sunday Mar 19	Module 1 Submit module 1 exercise
Mar 20 – Apr 2 Sunday Apr 2	Module 2 Submit module 2 exercise
Apr 3 – Apr 16 Sunday Apr 16	Module 3 Submit module 3 exercise
Thursday Apr 13 (5pm) Sunday Apr 30	Assignment 1 made available Submit assignment 1
May 1 – May 12 Friday May 12	Module 4 Submit module 4 exercise
May 13– May 24 Wednesday May 24	Module 5 Submit module 5 exercise
May 25 – Jun 4 Monday Jun 4	Module 6 Submit module 6 exercise
Fri Jun 2 (5pm) Sunday Jun 18	Assignment 2 made available Submit assignment 2

Acknowledgment of receipt of submitted exercises/assignments will be automatically generated using eLearning or via e-mail for e-mailed/mailed exercises. **However, it is the responsibility of the student to ensure that the assessment item (exercise or assignment) is received by the coordinator.**

Assessment

Assessment will include two written assignments worth 40% each. The assignments will be made available to students in the middle and near the end of the semester. The period between each assignment being made available and its due date includes three weekends (Australian time) and there will be a period of 14 days for each assignment during which no module material is being worked on concurrently.

In addition to the major assignments, students will be required to submit solutions to selected practical exercises (one from each module), worth a total of 20%.

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

Assignments

These must be submitted on or before the specified date. As a rule, NO variation of these dates will be allowed.

Please note: Requests for extension of the due date for an assignment must be made in advance of the starting date for that assignment. These requests must be made directly to the unit coordinator by email. The unit coordinator will reply by email with the decision as to whether an extension has been granted and the new due date.

Where a student is so incapacitated by a medical or other condition that he or she is unable to request an extension in advance, medical or other certification should explicitly note the severity of the disabling condition that precluded the advance request being made.

If no extension has been granted, late submissions will incur a penalty. Five percent of the earned mark for an assignment will be deducted for each day that an assignment is late, up to a maximum of 50%. It is not the intention of this late penalty policy to cause a student to fail the unit when otherwise they would have passed. If deductions for late assignments result in the final unit mark for a student being less than 50, when otherwise it would have been 50 or greater, the student's final mark will be exactly 50.

Exercises

There are a number of exercises for each module. These will be available on the eLearning website, they are currently not the same as the exercises in the textbook. One designated exercise from each module will be required to be submitted as part of the unit assessment. The date for submission of the exercises is given in the timetable section, below.

Please **name your document** with **your name and the module number**. Please submit word or rtf files wherever possible. If you submit documents in pdf format you will receive only brief comments on your solution.

Your solution to a submissible exercise should be posted in the Assignment section of PSI in eLearning. If you submit your exercise other than in eLearning, the marked exercise may be returned to you in the same way. We may not be able to post marked exercises onto eLearning if there is no submission on eLearning. The system will continue to say that the exercise has not been submitted.

The solution to the assessable exercise will be made available to you after submission. You will get immediate confirmation of the receipt of your submitted solution if you submit on eLearning.

PSI is a difficult course and has a heavy 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.

Submission of assessable material

Preferably, exercises and assignments should be submitted via eLearning. Alternatives are e-mail and post.

If the solution involves a substantial amount of calculus or algebra and you do not want to use the equation editor, then the solution can be mailed. If hand written solutions are sent, they should be neatly written and set out. Please note that WORD has an equation editor that allows you to write mathematical formulae.

Essentials for submission

- Your solution should be included in **one** WORD (or equivalent) document. Multiple files/documents will not be accepted. If you submit a Word document you will receive feedback on your submission, but if you submit a pdf the feedback is likely to be minimal.
- Do **not** send EXCEL spreadsheets.
- Do **not** send SAS or STATA code.
- 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.

Exercises and assignments must be submitted by **12 midnight (end of the day) on the due date** if submitting by email or on eLearning. If submitting by post within Australia, the assessment item should be sent by EXPRESS-Post and must be post-marked on or before the due date for submission (EXPRESS-Post deadline is posting by **6pm on submission day**). Please do not fax assignments or exercises.

The purpose of the course is not to teach statistical computing. There are some exercises that rely on the use of simulation to help understand the concepts being taught, and problems sometimes arise with running the software. Please do not leave these assessments until just before they are due – if you start early you will have time to ask questions and sort out the

problems. Do not allow difficulties with software to obscure the basis of the course, which is to understand the principles of statistical inference.

Important information that you must read on assessment and the declaration that must be submitted with assignments are on eLearning.

Further information regarding assessment submission procedures can be found at:

[BCA assessment guide for students](#).

Contact details

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

Liz Barnes

Email: liz.barnes@ctc.usyd.edu.au

Phone: 02 9562 5303

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: erica@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.