Study Guide

Principles of Statistical Inference (PSI)

Semester 1, 2015

Prepared by:
Professor Ian Marschner
NHMRC Clinical Trials Centre
University of Sydney

Copyright © NHMRC Clinical Trials Centre, 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 ....................................................................................................... 3
Unit Materials ................................................................................................................... 4
Outline of Unit Modules .................................................................................................. 5
Method of Delivery and Communication ................................................................. 6
  eLearning ..................................................................................................................... 6
  Email and phone communication ............................................................................. 6
Software ............................................................................................................................ 7
Reference Books ............................................................................................................. 7
Timetable and due dates ............................................................................................... 9
Assessment ....................................................................................................................... 10
  Assignments ............................................................................................................... 10
  Exercises .................................................................................................................... 10
  Submission of assessable material .......................................................................... 11
  Essentials for submission .......................................................................................... 11
Contact details ................................................................................................................. 12
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 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.

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 appendix of the PSI notes also contains a summary of basic terminology and concepts that are assumed knowledge for this unit. Please revise this material at the beginning of the unit.
Unit Materials

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

1. Module notes describing concepts and methods.
2. An extended example illustrating the concepts/methods introduced in the notes.
3. Practical exercises for students to complete.

Students should begin each module by reading through the module notes. The extended example should then be worked through in parallel with the exercises.

Some of the exercises require the student to perform computer simulations, and for these you may use either SAS or Stata (some notes and example code for both packages are provided with the modules notes and on eLearning).
Outline of Unit Modules

Module 1: Sampling variation
- Estimation
- Confidence intervals and inference
- Estimation – Concepts
- Probability models and parameters
- Estimates and estimators
- Delta method
- Point and interval estimation
- Bias and efficiency
- Consistency and asymptotic efficiency

Appendix : Review of basic probability
- Commonly used probability density functions

Module 2: Likelihood
- Likelihood function
- Sufficiency
- Nuisance parameters
- Approximate likelihood

Module 3: Estimation – Methods
- Maximum likelihood estimation
- Statistical information
- Properties of maximum likelihood estimation

Module 4: 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 5: Likelihood based methods
- Likelihood ratio tests
- Score tests
- Wald tests
- Relationship between the three tests
- Interval estimation based on the three tests

Module 6: Bayesian methods
- Basic concepts: subjective probability
- Bayes’ rule, prior and posterior distributions
- Conjugate and noninformative prior distributions
- Analysis of simple binomial & normal models
Method of Delivery and Communication

Liz Barnes and Lucy Davies are both biostatisticians at the NHMRC Clinical Trials Centre at the University of Sydney. Liz and Lucy will both respond to content-related questions in PSI and Liz is responsible for the administrative aspects of the course.

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 are given marks for their contribution to discussions on eLearning. While we will not use that practice here, we 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. To assist us in keeping track of e-mails, please use “PSI” in the subject line of your email when emailing the coordinators.

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 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 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.

Reference Books

The module notes and case studies form the primary material for this subject.

There is no single prescribed/compulsory text for the unit. However, the following book has been based on the material in this unit and is written by the same author as the module notes. It is recommended reading for the unit but not (yet) compulsory.

http://www.crcpress.com/product/isbn/9781482222234

A number of other books may provide background material. The following list is a selection of relevant texts, all of which are available at The University of Sydney. However, there are many other texts available which are suitable.


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
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.

<table>
<thead>
<tr>
<th>Date Range</th>
<th>Module/Task Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mar 2 – Mar 15</td>
<td>Module 1</td>
</tr>
<tr>
<td>Sunday Mar 15</td>
<td>Submit module 1 exercise</td>
</tr>
<tr>
<td>Mar 16 – Mar 29</td>
<td>Module 2</td>
</tr>
<tr>
<td>Sunday Mar 29</td>
<td>Submit module 2 exercise</td>
</tr>
<tr>
<td>Mar 30 – Apr 10</td>
<td>Module 3</td>
</tr>
<tr>
<td>Friday Apr 10</td>
<td>Submit module 3 exercise</td>
</tr>
<tr>
<td>Apr 10 after 5pm</td>
<td>Assignment 1 made available</td>
</tr>
<tr>
<td>Apr 10 – Apr 24</td>
<td>Work on assignment 1</td>
</tr>
<tr>
<td>Friday Apr 24</td>
<td>Submit assignment 1</td>
</tr>
<tr>
<td>Apr 25 – May 6</td>
<td>Module 4</td>
</tr>
<tr>
<td>Wednesday May 6</td>
<td>Submit module 4 exercise</td>
</tr>
<tr>
<td>May 7 – May 18</td>
<td>Module 5</td>
</tr>
<tr>
<td>Monday May 18</td>
<td>Submit module 5 exercise</td>
</tr>
<tr>
<td>May 19 – May 31st</td>
<td>Module 6</td>
</tr>
<tr>
<td>Sunday May 31st</td>
<td>Submit module 6 exercise</td>
</tr>
<tr>
<td>Jun 1 by 10 am</td>
<td>Assignment 2 made available</td>
</tr>
<tr>
<td>Jun 1 – June 14</td>
<td>Work on assignment 2</td>
</tr>
<tr>
<td>Sunday June 14</td>
<td>Submit assignment 2</td>
</tr>
</tbody>
</table>

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 2 written assignments worth 35% each. The assignments will be made available to students in the middle and near the end of the semester and are required to be completed within 2 weeks.

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 30%.

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%.

Exercises
There are a number of exercises at the end of each module. 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.

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 handwritten 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 or rtf you will receive feedback on your submission, but if you submit a pdf the feedback will 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.

Please note: 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 several exercises that rely on the use of simulation to help understand the concepts being taught, and problems sometimes arise with the code for these solutions. 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.
Contact details

For enquiries about this unit, contact the unit coordinators:

Liz Barnes
Email: liz.barnes@ctc.usyd.edu.au
Phone: 02 9562 5303

Lucy Davies
Email: lucy.davies@ctc.usyd.edu.au
Phone: 02 9562 5038

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.