Study Guide

Principles of Statistical Inference (PSI)

Semester 2, 2015

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

Copyright © The University of Sydney
# Study Guide
## Semester 2, 2015

## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background</td>
<td>2</td>
</tr>
<tr>
<td>Objectives</td>
<td>2</td>
</tr>
<tr>
<td>Assumed Knowledge</td>
<td>2</td>
</tr>
<tr>
<td>Unit Materials</td>
<td>3</td>
</tr>
<tr>
<td>Outline of Unit Modules</td>
<td>3</td>
</tr>
<tr>
<td>Method of Delivery and Communication</td>
<td>4</td>
</tr>
<tr>
<td>eLearning</td>
<td>4</td>
</tr>
<tr>
<td>Email and phone</td>
<td>5</td>
</tr>
<tr>
<td>Reference Books</td>
<td>5</td>
</tr>
<tr>
<td>Software</td>
<td>6</td>
</tr>
<tr>
<td>Timetable</td>
<td>7</td>
</tr>
<tr>
<td>Assessment</td>
<td>8</td>
</tr>
<tr>
<td>Extensions</td>
<td>9</td>
</tr>
<tr>
<td>Late penalty</td>
<td>9</td>
</tr>
<tr>
<td>Contact details</td>
<td>10</td>
</tr>
</tbody>
</table>
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. 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.

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.

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. Appendix 1 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 approximately 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 within the modules notes and on the eLearning site).

Outline of Unit Modules

Module 1: Estimators and estimates
- Parameters, estimators and estimates
- Bias and efficiency
- Consistency and asymptotic efficiency
- Delta method
- Confidence intervals

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

Module 3: Estimation methods
- Maximum likelihood estimation
- Fisher 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 non-informative prior distributions
- Analysis of simple binomial & normal models

Appendix 1: Review of basic probability
Appendix 2: Commonly used probability distributions

Method of Delivery and Communication

eLearning
We will use the BCA eLearning site (also known as Learning Management System (LMS) or Blackboard) as the main means of communication. An eLearning Guide, which gives basic information on how to use online eLearning is available from the Student Resources page http://www.bca.edu.au/currentstudents.html on the BCA website.

The eLearning website should be the primary mode for asking questions, via the Discussion facility. In later units, students are given marks for their contribution to discussions. While we will not use that practice here, we nevertheless encourage you to use the Discussion facility 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 coordinators will monitor the discussions and contribute when needed.

Every effort will be made by the instructors to intervene in discussions, if necessary, in a timely manner between Monday and Friday (within 48 hours from posting, depending on the nature of the problem). Please note that instructors will not be available to reply to email or questions on eLearning during the weekends.

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.

The eLearning website 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 eLearning 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.

Reference Books

The module notes form the primary material for this subject.

There is no single prescribed/compulsory text for the unit, but a number of reference books are recommended as background material. The following list is a selection of relevant recommended 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
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.

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.
**Timetable**

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.**

<table>
<thead>
<tr>
<th>Week</th>
<th>Topic</th>
<th>Tuesday date</th>
<th>Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Module 1</td>
<td>28 July</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>4 August</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Module 2</td>
<td>11 August</td>
<td>Module 1 exercise due</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>18 August</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Module 3</td>
<td>25 August</td>
<td>Module 2 exercise due</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>1 September</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Module 4</td>
<td>8 September</td>
<td>Module 3 exercise due</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Assignment 1 handed out</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>15 September</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>22 September</td>
<td>Assignment 1 due</td>
</tr>
<tr>
<td>10</td>
<td>Mid-semester Break</td>
<td>30 September</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Module 5</td>
<td>6 October</td>
<td>Module 4 exercise due</td>
</tr>
<tr>
<td>12</td>
<td>Module 6</td>
<td>20 October</td>
<td>Module 5 exercise due</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>27 October</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>3 November</td>
<td>Module 6 exercise due</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>10 November</td>
<td>Assignment 2 handed out</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>17 November</td>
<td>Assignment 2 due</td>
</tr>
</tbody>
</table>
Assessment

The assessment for this unit is based entirely on assignments and submitted exercises. There is no examination. There are two assignments worth 35% each. In addition, one designated exercise from each module will be required to be submitted as part of the unit assessment (worth 5% each).

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

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Due Date</th>
<th>% marks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Module 1 exercise</td>
<td>11 August</td>
<td>5</td>
</tr>
<tr>
<td>Module 2 exercise</td>
<td>25 August</td>
<td>5</td>
</tr>
<tr>
<td>Module 3 exercise</td>
<td>6 September</td>
<td>5</td>
</tr>
<tr>
<td>Assignment 1</td>
<td>22 September</td>
<td>35</td>
</tr>
<tr>
<td>Module 4 exercise</td>
<td>6 October</td>
<td>5</td>
</tr>
<tr>
<td>Module 5 exercise</td>
<td>20 October</td>
<td>5</td>
</tr>
<tr>
<td>Module 6 exercise</td>
<td>3 November</td>
<td>5</td>
</tr>
<tr>
<td>Assignment 2</td>
<td>17 November</td>
<td>35</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>100</strong></td>
<td></td>
</tr>
</tbody>
</table>

Module exercises are located at the end of each module. Assignments will be posted online. All assessments are due on **Tuesday by 11:59pm**.

Assessed exercises and assignments should be submitted on the eLearning website. Please consult the BCA Assessment Guide for details about submitting your assignments, and guidelines for written work. In particular:

- Your solution should be included in **one** WORD or PDF document. Multiple files/documents will **not** be accepted.
- **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 use the equation editor in WORD for writing mathematical formulae. Marks will be deducted if incorrect notation is used.
- **Please name your WORD document with your name and the module or assignment number.** When the exercises are downloaded from eLearning having 60 exercises all named *module3.doc* can be problematic.
All material submitted for assessment must be entirely your own work. See the section on Academic Dishonesty and Plagiarism of the BCA Assessment Guide. Students are reminded that plagiarism is not acceptable. Please ensure that you understand what plagiarism means and that you have read and understood the policy associated with the university in which you are enrolled. When submitting material for assessment, you will be required to complete a declaration, in the form of an either/or test question, certifying that you have done this.

We strongly recommend that you keep a copy of your assignments.

Acknowlgedgment of receipt of submitted exercises/assignments will be automatically generated in eLearning or via e-mail for e-mailed exercises. It is the responsibility of the student to ensure that the assessment item is received by the unit coordinator by the due date.

Extensions
Request for an extension of the due date for an assessment must be made in advance of the due date. This request 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 given 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.

Late penalty
Five percent (5%) of the total marks for an assessment will be deducted for each day that an assessment is late, if no extension has been given. There are administrative reasons for having these penalties. The markers will want to mark all the submissions for an assessment at the same time. This is not only more efficient but it is more likely that the marks allocated will be consistent. Another reason is that PSI is a difficult course with a heavy workload; hence, if you fall behind it will be difficult to catch up.
Contact details

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

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 9036 5337
fax: 02 9351 5049

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

Professor Judy Simpson
Sydney School of Public Health
Edward Ford Building (A27), University of Sydney, NSW 2006
email: judy.simpson@sydney.edu.au
phone: 02 9351 4369
fax: 02 9351 5049

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.