Data Management and Statistical Computing (DMC)

Study Guide
(course outline)

Semester 1, 2016

Centre for Clinical Epidemiology and Biostatistics;
School of Medicine and Public Health;
Faculty of Health;
The University of Newcastle
## Contents

1 Preface .......................................................................................................................................................... 1  
2 Course Outline .............................................................................................................................................. 2  
   2.1 Letter of Introduction .............................................................................................................................. 2  
   2.2 Course Objectives ......................................................................................................................................... 3  
   2.3 Course Content .......................................................................................................................................... 3  
      2.3.1 Module Topics ..................................................................................................................................... 3  
   2.4 Online Learning ......................................................................................................................................... 4  
   2.5 Using Statistical Programs in Biostatistics ............................................................................................... 4  
      2.5.1 Course Schedule ............................................................................................................................. 5  
   2.6 Learning Resources ..................................................................................................................................... 5  
      Textbook(s) ................................................................................................................................................. 5  
      Readings ...................................................................................................................................................... 5  
      Software ....................................................................................................................................................... 6  
   2.7 Assessment Items ...................................................................................................................................... 6  
   2.8 Assignment Submission ............................................................................................................................. 6  
3 Data Management and Statistical Computing: Course Readings ................................................................ 7
1 Preface

Course Title: Data Management and Statistical Computing
Unit Value: 10
Term: Semester 1, 2016
Mode of Delivery: Distance Learning, Online
Managing Campus: Callaghan, University of Newcastle
CRICOS Provider No: 00109J (The University of Newcastle)
Assumed Knowledge: Mathematical Background for Biostatistics
Course Requisites: Nil
Contact Hours: 1. Self-Directed Learning for 6 hours per week for full term. 2. Online Discussion Group for 2 hours per week for full term.
Contact hours are an indication only.
Workload: Students are required to spend on average 120 – 140 hours of effort or total load (contact and non-contact including assessment) per term per 10 unit value, whatever the discipline or the mode of delivery of the Course
Core Component: This course is a Compulsory Program Component of the following Programs:
1. Graduate Diploma in Medical Statistics
2. Master of Medical Statistics
Course Coordinator: Professor Patrick McElduff
Email: Patrick.McElduff@newcastle.edu.au
School Office: School of Medicine and Public Health, Faculty of Health
Email: smph-pg@newcastle.edu.au
BCA Contact: If you have trouble contacting the course coordinators or have any other queries, please contact:
Erica Jobling
Executive Officer
Biostatistics Collaboration of Australia
Email: erica@ctc.usyd.edu.au
Phone +61 2 9562 5076
Fax: +61 2 9565 1863
GradSchool: Email: postgradstudy@newcastle.edu.au
Phone +61 2 4921 8856
Academic Misconduct: Academic integrity, honesty, and a respect for knowledge, truth and ethical practices are fundamental to the business of the University. These principles are at the core of all academic
endeavour in teaching, learning and research. Dishonest practices contravene academic values, compromise the integrity of research and devalue the quality of learning. To preserve the quality of learning for the individual and others, the University may impose severe sanctions on activities that undermine academic integrity.

**Alteration of this Course Outline**

No alteration to the approved course outline will be permitted without Head of School approval. Students will be notified of any change and the revised Course Outline will be provided to students in the same manner as the original.

## 2 Course Outline

### 2.1 Letter of Introduction

Welcome to the Data Management and Statistical Computing (DMC) component of your program. This course will develop statistical/computing skills that can be applied to data management and analysis of health and medical studies.

For some of you, this may be one of the first courses you are undertaking as part of your BCA program, so welcome to the BCA program. My name is Patrick McElduff and I am the DMC course co-ordinator for Semester 1.

Blackboard is the elearning system for BCA courses. There is an email facility on Blackboard, but please do NOT use this. You should post any content related questions on Blackboard. If you have any administrative questions, please direct these to my email.

Hopefully, there will be plenty of postings on the discussion board. Please don't feel shy about either posting a comment or question (if you have a query there is a high likelihood that others will have the same concern/query) or responding to someone else's comments or questions. I will be keeping a close eye on the discussion board, and contributing to this, but will not necessarily respond every day. Ideally, I would like to see students respond to each other first. I will aim to ensure that posted items are responded to (by other students or me) within 2 working days.

As a good way to get to know each other, a directory will be set up where you can, if you wish, tell others who you are and what your background is.

The module notes and datasets are all posted on Blackboard, so you should be able to download and print them if you wish. However copies of articles cannot be posted on Blackboard as this breaches copyright.

The Course Information booklet includes the timetable for assessments and other important information such as an Introduction to Online Learning, to help get you started in Blackboard; an Assessment Guide, with information on how to structure and submit assessments and link to university plagiarism policies; and an assessment cover sheet. These items are available on Blackboard.
Assessments will be posted on Blackboard, and an announcement indicating this will be posted. You should regularly check for announcements as well as discussions. Additional datasets and/or other files will generally be required for assessments and these will be posted on Blackboard along with the assessments. Assessments should be submitted electronically through Blackboard, as it is easier to keep track of submission this way. If you submit your assessment via email directly to me (not recommended!), you will need to check that it has been received.

It is very important that you organise the software required for this course as soon as possible – it may take some time to purchase and obtain copies of SAS and Stata, and delays may impact on your ability to complete this course.

Please do not hesitate to contact any of the DMC staff should you wish to ask questions or make any comments on any aspect of the course. Any students with administrative difficulties or concerns should contact the relevant Education Office staff members.

I sincerely hope that you enjoy the Data Management and Statistical Computing course. All the best with your studies!

Dr Patrick McElduff
Course Coordinator
Data Management and Statistical Computing

2.2 Course Objectives

Course objectives are to provide students with skills to undertake moderate to high level data management, manipulation and analysis. Specific objectives are for students to:

1. Gain experience in data manipulation and management using two major statistical software packages (Stata and SAS)
2. Learn how to display and summarise data using statistical software
3. Become familiar with the checking and cleaning of data
4. Learn how to link files through use of unique and non-unique identifiers
5. Acquire fundamental programming skills for efficient use of software packages
6. Learn key principles regarding confidentiality and privacy in data storage, management and analysis

2.3 Course Content

2.3.1 Module Topics

The topics to be covered are:

- Module 1 – Stata and SAS: The basics (importing and exporting data, recoding data, formatting data, labelling variable names and data values; using dates, data display and summary presentation)
- Module 2 – Stata and SAS: graphs, data management and statistical quality assurance methods (including advanced graphics to produce publication-quality graphs)
Module 3 – Data management using Stata and SAS (using functions to generate new variables, appending, merging, transposing longitudinal data; programming skills for efficient and reproducible use of these packages, including loops, arguments and programs/macros)

The suggested learning process each week is for the student to read the relevant material, then attempt the exercises provided within the module.

Students with questions regarding the contents of the course are encouraged to post them to the discussion board facility on Blackboard. Other students are then encouraged to answer these questions, offer suggestions and raise related questions thereby creating an online tutorial via Blackboard. The course co-ordinator will oversee all discussions, provide answers to questions, make suggestions and raise other questions.

2.4 Online Learning

The course is taught as an online course. For information on eLearning see, “BCA Introduction to Online Learning: Getting Started in eLearning” which can be accessed from the BCA student resource page.

2.5 Using Statistical Programs in Biostatistics

This course provides an introduction to SAS and Stata, and covers only a very small component of these programs. However, you will need to substantially expand your knowledge to be an effective biostatistician. Part of what we intend to do with this course is give you a basis that you can build on, and the ability and confidence to be able to ‘lead yourself’ to advance your skills.

Thus, the assessments may require you to go beyond the material in the notes and should extend you.
2.5.1 Course Schedule

Table 1 Study Timetable

<table>
<thead>
<tr>
<th>Month</th>
<th>Date (Monday)</th>
<th>Academic Week</th>
<th>Modules</th>
<th>Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feb</td>
<td>29</td>
<td>1</td>
<td>Module 1</td>
<td></td>
</tr>
<tr>
<td>Mar</td>
<td>7</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25th March-1st April</td>
<td>4 5</td>
<td></td>
<td>1st Assignment due Monday 4 April</td>
<td></td>
</tr>
<tr>
<td>April</td>
<td>11</td>
<td>6</td>
<td>Module 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>May</td>
<td>2</td>
<td>9</td>
<td>Module 3</td>
<td>2nd Assignment due Monday 2 May</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>June</td>
<td>6</td>
<td></td>
<td>Exam Period Commences</td>
<td>3rd Assignment due Monday 6 June</td>
</tr>
</tbody>
</table>

2.6 Learning Resources

Textbook(s)

You are required to purchase (or borrow from a university library) the following textbooks:


Readings

Readings are compulsory. All of the articles are available through ScienceDirect Journals. You should be able to access them through your University Library’s Journal Online Database or request copies via your library’s resource request system.
Software

You are required to have access to the following software packages:

- **STATA** – version 12 or later
- **SAS** – version 9.2 or later

You should make it a priority to organise access to these programs. It can take some time to obtain a copy of these programs, particularly SAS, so please do not delay and do this as soon as possible.

You will need to have access to Stata and SAS for this course. Information on purchase of software and texts is provided on the BCA website: [http://www.bca.edu.au/currentstudents.html](http://www.bca.edu.au/currentstudents.html)

We would like to stress again the importance of obtaining copies of the software required as soon as possible.

Stata version 14 is now available. New students should purchase version 14 and students who have been enrolled in previous years may have an earlier. The notes will be based on version 12. A summary of the main differences between versions can be obtained from the survey design website: [http://www.survey-design.com.au/](http://www.survey-design.com.au/) or the Stata website.

### 2.7 Assessment Items

Assessment for the course will be based on three Assignments.

<table>
<thead>
<tr>
<th>Assessment Item</th>
<th>Weighting in Course</th>
<th>Due Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignment 1</td>
<td>30%</td>
<td>11.59pm on Monday 4 April</td>
</tr>
<tr>
<td>Assignment 2</td>
<td>35%</td>
<td>11.59pm on Monday 2 May</td>
</tr>
<tr>
<td>Assignment 3</td>
<td>35%</td>
<td>11.59pm on Monday 6 June</td>
</tr>
</tbody>
</table>

Assignments 1, 2 and 3 will be placed on Blackboard **two (2) weeks** prior to the due date. All assignments are to be submitted on Blackboard.

### 2.8 Assignment Submission

For more information and guidelines for written work, submission of assignments and exams, and BCA policies and procedures, click on [BCA Assessment Guide](http://www.bca.edu.au/currentstudents.html).

**Penalty for assignments submitted after the due date**

If no extension has been given, 5% of the earned mark for an assignment will be deducted for each day that an assignment is late, up to a maximum of 50%. NOTE: 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.
You should be able to access the course readings through the library at your home university (see student resource page for links)


Hosking JD, Newhouse BA, Bagniewska A, Hawkins B. Data collection and transcription. Controlled Clinical Trials 1995; 16(2) Supp1: 66S-103S.


