



Study Guide

Data Management and Statistical Computing (DMC)

Semester 1, 2019

Prepared by: Michael Waller
School of Public Health
The University of Queensland

Murthy N Mittinty
School of Public Health
The University of Adelaide

Copyright © The University of Queensland, The University of Adelaide, The University of Melbourne



CONTENTS

Instructor contact details	3
Other Contacts	3
Welcome Letter	4
Background	5
Course Content.....	5
Workload requirements.....	6
Recommended approaches to study	6
Method of communication with coordinator.....	6
Text Books	7
Readings.....	7
Software.....	7
Module Descriptions.....	7
Course Timetable	8
Assessment	8
Submission of assessments and academic honesty policy.....	9
Extensions	9
Penalties for Late Submission	9
Assignment Cover Sheet	10
Feedback	10
Changes to DMC since last delivery, including changes in response to student evaluation	10
Acknowledgements.....	10

Instructor contact details

Dr Murthy Mittinty

School of Public Health, The University of Adelaide, AHMS Building
57 North Terrace
Adelaide 5000

Email: murthy.mittinty@adelaide.edu.au

Other Contacts

If you have trouble contacting the academic coordinator/academic staff, or have any other queries, please contact:

Erica Jobling /Emily Higginson

BCA Coordinating Office
Biostatistics Collaboration of Australia BCA c/o NHMRC Clinical Trials Centre Locked Bag 77
Camperdown NSW 1450

Phone: 02 9562 5076 Email: bca@ctc.usyd.edu.au

Welcome Letter

Welcome to Data Management and Statistical Computing (DMC). In this unit we will develop statistical computing skills essential for managing and analysing data in health and medicine. This course provides an introduction to R and Stata, with the aim of giving you a foundation to build upon in your further studies and in your biostatistical career. From this semester onwards we no longer introduce SAS and have replaced it with the very popular (and freely available) R. Modules 2A, 2B and 3 have been updated with R code. This unit is delivered through the eLearning site at the University of Sydney. All course content other than readings (discussed below) will be uploaded to eLearning, including assignments and supplementary material. Discussions of material will take place on the Discussion Board. There is currently an Introductions thread on Discussion Board; please use this thread to introduce yourself to the rest of the class. This unit requires access to two statistical software packages: R and Stata (detailed shortly). You should organise access to these as soon as possible. Since this is the first time we have introduced R, we would appreciate as much feedback as possible for continuing improvement of the unit material for future deliveries.

If you have any questions or issues, please contact me by email at the address above. I hope you enjoy the course!

Murthy Mittinty
March 2019

Background

This course aims to provide students with skills to undertake moderate to high level data management, manipulation, and analysis. On completion of this unit, students should:

1. Be able to undertake data manipulation and management using two major statistical software packages (Stata and R);
2. Be able to appropriately display and summarise data using statistical software;
3. Understand how to check and clean data;
4. Be able to link data files through unique and non-unique identifiers;
5. Have fundamental programming skills for efficient use of statistical software;
6. Understand key principles of confidentiality and privacy in data storage, management and analysis.

Course Content

This course consists of three modules:

- Module 1: The basics. Importing and exporting data; recoding and formatting data; labelling variables and values; use of date data, displaying and summarising data.
- Module 2: Graphs, Data management and Statistical Quality Assurance Methods. Includes advanced graphics for production of publication-quality graphs.
- Module 3: Data Management. Using functions to generate new variables; appending, merging and transposing data; programming skills including loops, arguments and programs/macros.

Each module requires approximately 4 weeks of study; the final week of semester will be left for revision, or to cover other issues which arise during the course.

Course material consists of (a) the notes which are provided for each module, (b) the text books and other required reading, and (c) further notes, code and data files which will be provided on eLearning.

Study materials for all Modules are downloadable from the eLearning (Canvas) unit site. Assignments and supplementary material, such as data sets will be posted to the unit site. Please note that we are not able to post copies of copyright material (journal articles and book extracts)—for these you will have to rely on resources from your home university's library.

Workload requirements

The expected workload for this unit is 10-12 hours per week on average, consisting of guided readings, discussion posts, independent study and completion of assessment task.

Recommended approaches to study

Students should work through each module systematically, following the module notes and any readings referred to, and working through the accompanying exercises. *You will learn a lot more efficiently if you tackle the exercises systematically as you work through the notes.* 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.

Outline 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 attack the exercises independently (or via the eLearning site), and yet not make you wait too long to see the sketch solutions.

Method of communication with coordinator

Questions relating to unit content can be emailed to the coordinator. Coordinators will be able to answer questions relating to the module notes and practical exercises, and to address any other issues that require clarification. Please note that it might take a day or two to respond to your queries (possibly longer during breaks and over weekends!). Please post content-related questions to the discussion board on canvas.

We strongly recommend that you post content-related questions to the Discussions tool in the (DMC) area of BCA's eLearning site. In 2019 we are using the Learning Management 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.

Text Books

You should have access to the following textbooks:

Jull S, Frydenberg M. An Introduction to Stata for Health Researchers, 4th ed. Stata Press, 2014.

Dalgaard, P. Introductory Statistics with R, 2nd ed. Springer, 2008.

I recommend you check your University Libraries to see if ebook (Full Text Online) versions are available. If you have any issues accessing these text books please contact me.

Readings

In addition to the text books, various other materials are set as required or supplementary readings in each module. These cannot be uploaded to eLearning but you should be able to access these articles through your university's library; further assistance in accessing readings will be given during the course if necessary.

Software

You should have access to the following software packages:

- Stata version 12 or later (the latest version is v15)
- R version R64 3.4.2 or later

If you have not yet organised access to these packages, you should do so as soon as possible. This is a practical course which requires regular use of the relevant software; delays in gaining access to these packages may impact your ability to complete the course.

Information on how to download R, and access Stata can be found in the [BCA Textbook and Software Guide](#).

Module Descriptions

As described above, there are 3 modules in this course here we outline the timetable and assessment description.

Each module is scheduled to begin on a Monday and conclude on the Sunday of the following week. The due date for submission of required assignments from each module is 11.59 pm on the day it is described to be due as indicated below in the table.

Course Timetable

Semester 1, 2019 will commence on Monday 4th March.

Week	Week Commencing	Module	Assessment
1	Monday 4 th March	1	
2	Monday 11 th March	1	Assignment 1 Available Friday 15 th March
3	Monday 18 th March	1	
4	Monday 25 th March	1	
5	Monday 1 st April	2	Assignment 1 Due Monday 1st April
6	Monday 8 th April	2	
7	Monday 15 th April	2	Assignment 2 Available Friday 19 th April
Monday 22 nd April to Sunday 28 th April - Mid Semester Break			
8	Monday 29 th April	2	
9	Monday 6 th May	3	Assignment 2 Due Monday 6th May
10	Monday 13 th May	3	Assignment 3 Available Wednesday 15 th May
11	Monday 20 th May	3	
12	Monday 27 th May	3	
			Assignment 3 Due Saturday 1st June

Assessment

The assessment for this unit consists of three assignments:

- Assignment 1 will cover Module 1, and is worth 30% of the overall course mark. It is due before midnight (EST) on Monday 1st April 2019.
- Assignment 2 will cover Module 2, and is worth 35% of the overall course mark. It is due before midnight (EST) on Monday 6th May 2019.
- Assignment 3 will cover Module 3, as well as Modules 1 and 2, and is worth 35% of the overall course mark. It is due before midnight (EST) on Saturday 1st June 2019.

All assignments will be posted on eLearning 2.5 weeks before the due date. Individual feedback will be provided to each student; model solutions will also be provided once all marked assignments have been returned. Summary statistics on results for the entire class will also be provided.

Assignments should be submitted via the assignment submission tool on eLearning; if you experience difficulties with this submission method, assignments can be submitted via email.

Submission of assessments and academic honesty policy

All assessment material must be submitted via canvas unless otherwise advised. The use of Turnitin for submitting assessment items has been instigated within unit sites. For more details 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.

Extensions

For various reasons, you may sometimes experience difficulties in getting your assignments submitted on the due date. Requests for an extension for an assignment must be made **in advance of the due date for that assignment**. The normal grounds for an extension being granted are bereavement, personal illness or illness in a family member requiring you to exercise a significant career role.

These requests must be made directly to Murthy Mittinty by email, and should include appropriate documentation (e.g. medical certificate). The time and date of the request will be noted, and a reply sent by email with the decision as to whether an extension has been granted and, if so, stating the length of the extension.

Length of extension: Extensions granted by Unit Coordinators will normally be no longer than three days. Extensions longer than 3 days must be approved by your home University.

Penalties for Late Submission

BCA policy states: Unless otherwise stated, a student can submit an assessment up to 10 days after the due date. A late penalty of 5% per day will be applied (including weekends and public holidays). The maximum penalty which can be applied is a reduction to 50% of the total assessment mark.

Submissions after the solutions have been posted on eLearning will not be awarded any marks.

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.

Assignment Cover Sheet

Where assignment work is submitted online using the Assignment tool in canvas you will be able to indicate your compliance with the plagiarism guidelines and policy by electronic means. In this case, you **do not** need to complete the DMC 2019 Assignment Cover Sheet.

If you submit work by another method, then you **do** need to complete the DMC 2019 Assignment Cover Sheet, in which you will be asked to certify that the submission is your own work and that you have read the policy of the university at which you are enrolled. The cover sheet can also be downloaded from canvas. If you are posting your submission, please include the signed cover sheet in the envelope.

If you are submitting via email, please scan the signed cover sheet and submit this with your assignment, or fax the signed cover sheet to the number specified on the sheet.

Please refer to the [BCA Assessment Guide](#) for further information.

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 Canvas

Your feedback to us:

One of the formal ways students have to 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.

Changes to DMC since last delivery, including changes in response to student evaluation

DMC was last delivered in Semester 2 2018. There has been a major change since that delivery in the form of introducing R and removing SAS. Some minor changes include typos and minor edits for greater clarification of the text.

Acknowledgements

We thank The University of Melbourne for sharing its learning modules which aided in preparing the R module for this semester.