BIO STATISTICS COLLABORATION OF AUSTRALIA

Study Guide

Data Management and Statistical Computing (DMC)

Semester 2, 2021

Prepared by:

Jennie Louise
School of Public Health
The University of Adelaide

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

THE UNIVERSITY of ADELAIDE

Contents

Data Management and Statistical Computing (DMC)
Semester 2, 2021

Instructor contact details

Instructor David Fitzgerald

School of Public Health
Faculty of Medicine
The University of Queensland
266 Herston Rd, Herston, QLD 4006

(07) 3365 5514

d.fitzgerald@sph.uq.edu.au

Other Contacts
If you have trouble contacting the academic coordinator/academic staff, or have any other queries, please contact:
Karolina Kulczynska-Le Breton /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. 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 we have recently introduced R into the course in 2019, 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!

David Fitzgerald
July 2021

Background and Unit Summary

The aim of this unit is to provide students with the knowledge and skills required to undertake moderate to high-level data manipulation and management in preparation for statistical analysis of data typically arising in health and medical research.

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 tasks.

Prerequisites

None

Co-requisites

None

Learning Outcomes

At the 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.

Unit content

The unit is divided into 3 modules, summarised in more detail below. Each module will involve approximately 4 weeks of study and generally includes the following material:

• 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: More Advanced Statistical Computing: Using functions to generate new variables; appending, merging and transposing data; programming skills including loops, arguments and programs/macros.

Study materials for all Modules are downloadable from the eLearning unit site. Assignments and supplementary material, such as datasets 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 the hard copy mail-out or resources from your home university's library.

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.

Method of communication with coordinator(s)

Questions about administrative aspects or course content can be emailed to the coordinator, and when doing so please use "DMC:" in the Subject line of your email to assist in keeping track of our email messages. Coordinator/s will be available to answer questions related to the module notes and practical exercises, and to address any other issues that require clarification. However, please note that instructors are not necessarily available every day of the week and you should expect that it may take a day or so to respond to questions (possibly longer over weekends and during breaks!).

We strongly recommend that you post content-related questions to the Discussions tool in the DMC area of BCA's eLearning site. In 2021 we are using the Learning Management system hosted by the University of Sydney. You may be familiar with the 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.

Module descriptions

As described above, there are 3 modules in this course; each module has been divided into Part A and Part B, each scheduled over a fortnight. Each module sub-section 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 11:59pm (Australian Eastern Standard Time) on the due date.

Unit schedule

Semester 2, 2021 starts on Monday 26 July

Week	Week commencing	Module	Assessment
1	Monday 26th July	1A	
2	Monday 2 August	1A	Assignment 1 available Friday 6th August
3	Monday 9 August	1B	
4	Monday 16 August	1B	
5	Monday 23 August	2A	**Assignment 1 Due Monday 23 August**
6	Monday 30 August	2A	Assignment 2 available Friday 3rd September
7	Monday 6 September	2B	
8	Monday 13 September	2B	
9	Monday 20 September	3A	
Mid Semester Break			
10	Monday 4 October	3A	**Assignment 2 Due Monday 4th October** Assignment 3 Available Friday 8th October
11	Monday 11 October	3B	
12	Monday 18 October	3B	
13	Monday 25 October		**Assignment 3 Due Monday 25th October**

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 11:59pm (AEST) on Monday 23rd August 2021.
- Assignment 2 will cover Module 2, and is worth 35% of the overall course mark. It is due before 11:59pm (AEST) on Monday 4th October 2021.
- 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 11:59pm (AEST) on Monday 25th October 2021.

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.

See the BCA Assessment Guide document for specific guidelines on acceptable standards for assessable work.

The instructors will generally avoid answering questions relating directly to the assessable material until after it has been submitted, but we encourage students to discuss the relevant parts of the notes among themselves, via eLearning. However, **explicit solutions to assessable exercises should not be posted for others to use**, and each student's submitted work must be clearly their own, with anything derived from other students' discussion contributions clearly attributed to the source.

Submission of assessments and academic honesty policy

You should submit all your assessment material via eLearning unless otherwise advised. The use of Turnitin for submitting assessment items has been instigated within unit sites. For more detail 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.

Late submission of assessments and extension procedure

The standard BCA policy for late penalties for submitted work is a 5% deduction from the earned mark for each day the assessment is late, up to a maximum of 10 days (including weekends and public holidays). Extensions are possible, but these need to be applied for (by email) as early as possible. The Unit Coordinator is not able to approve extensions beyond three days; for extensions beyond three days you need to apply to your home university, using their standard procedures.

Learning resources

Text Books

It is recommended that you have access to the following textbooks:
Jull S, Frydenberg M. An Introduction to Stata for Health Researchers, 4th ed. Stata Press, 2014.

Wickham H, Grolemund G. R for Data Science. O'Reilly 2017. Dalgaard, P. (available online https://r4ds.had.co.nz/)

Your University Library may have an ebook (Full Text Online) version of the Juul text; the Wickham text is freely available at the web link provided. If you have any issues accessing these texts please contact me.

Readings

In addition to the text books, various other materials may be set as required or supplementary readings in each module. These cannot be uploaded to eLearning but you should be able to access them 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 v16)
- R version R64 3.4.2 or later (the latest version is 4.0.2)
- RStudio version 1.3 or later (the latest version is 1.4)

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 RStudio, and access Stata can be found in the BCA Textbook and Software Guide.

Feedback

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 1 2021. In Semester 1, 2020, the R notes were redeveloped by Dr Jennie Louise, and further changes, including incorporation of online tutorials and video content, was implemented in Semester 2, 2021. Further changes have been implemented in Semester 1, 2021, including additional modification of the course notes and incorporation of optional supplementary content.