



Study Guide

Health Indicators and Health Surveys (HIS)

Semester 1, 2018

Prepared by:

Armando Teixeira-Pinto

School of Public Health, The University of Sydney

Copyright © School of Public Health, The University of Sydney



THE UNIVERSITY OF
SYDNEY

Health Indicators and Health Surveys (HIS)

Study Guide

Contents

Unit Overview	3
Unit Objectives	3
Assumed Knowledge	3
Unit Structure	4
Timetable	5
Module Objectives	6
Methods of Communication	11
Textbook	13
Software	14
Assessment	14
Contact details	16

Unit Overview

Practising biostatisticians need to be aware of the main sources of routinely collected health and demographic data, such as censuses and national health surveys, and how to make valid inferences from and comparisons with these data, using standard demographic techniques. They should also be able to design a health survey to collect primary data, obtain a random sample of the target population efficiently, and analyse the data taking the sampling strategy into account.

In this unit you will be introduced to a variety of sources of routinely collected health-related data and how these data are used to derive population measures of fertility, mortality and morbidity, and to measure health service utilisation, disease registration and reporting. You will learn to use quantitative demographic methods of direct and indirect age standardisation, and calculation of life expectancy by life table techniques to obtain valid comparisons between different population groups, and to examine health differentials.

You will also learn to develop, design and deliver a valid and reliable health questionnaire. The use of focus groups and other qualitative methods is described, as are standard instruments used in health surveys, coding, validity and reliability of measures, advantages and disadvantages of various modes of data collection. Finally, you will learn to design and implement an efficient sampling strategy, and to analyse and interpret the data, taking into account design issues, such as stratification, clustering and weighting.

Unit Objectives

At the completion of this unit you should be able to:

1. derive and compare population measures of mortality, illness, fertility and survival, using basic demographic tools such as life tables and age standardisation;
2. access the main sources of routinely collected health data and choose the appropriate one, taking into account their advantages and disadvantages;
3. design a valid and reliable health survey to collect primary data, design an efficient sampling strategy to obtain a random sample of the target population, and choose the most appropriate mode of delivery;
4. estimate means, totals and proportions from survey data, taking into account the sample design, and analyse, interpret and present these results.

Assumed Knowledge

Students should already be familiar with the Normal and binomial distributions, their mean and variance, estimation of means and proportions with confidence intervals, and the comparison of means and proportions between two groups using hypothesis tests

(i.e. t-tests and chi-squared tests for 2×2 tables). If you need to revise any of these concepts the following resources may be useful:

(HIS textbook) Levy PS and Lemeshow S. *Sampling of Populations: Methods and Applications*, Chapter 2

Bland JM. *An Introduction to Medical Statistics*, Chapters 1, 3-9

Swinscow, *Statistics at Square One*, Ch 1-4 <http://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one>

SurfStat Australia <http://surfstat.anu.edu.au/surfstat-home/surfstat-main.html>

See also page 14 for details of assumed knowledge about Excel and either SAS or Stata.

Unit Structure

This unit is offered throughout Australia through the Biostatistics Collaboration of Australia. It is available in distance learning mode only, to students enrolled in postgraduate degrees in biostatistics coordinated by the BCA.

The unit consists of 4 modules comprising 3 or 4 topics, each designed to take 1 week to complete (see timetable below). Each module comprises a set of notes or a study guide with readings, some exercises and an assignment. Modules 3 and 4 also have several videos about the topics covered and available in the website.

Timetable

	Week begins	Feb 26	Mar 5	Mar 12	Mar 19	Mar 26	Apr 2 BREAK	Apr 9	Apr 16	Apr 23	Apr 30	May 7	May 14	May 21	May 28	June 4	
Module																	
	Wk	1	2	3	4	5	E A S T E R	6	7	8	9	10	11	12	13		
1		1	2	3													
2					4	5		6									
3									7	8	9						
4												10	11	12	13		
Assignment released			1		2						3				4		
Assignment due					1				2 (1A)	Feedback to classmates	2 (rest)		3				4

Numbers in the shaded cells correspond to topic numbers, as shown in the table below.

The modules are:

Module	Week	Topic
1	1	1. Introduction to health indicators, mortality and fertility
	2	2. Standardisation
	3	3. Life tables
2	4	4. Routinely collected data and measurement of morbidity
	5	5. Design of sample surveys
	6	6. Questionnaire administration, testing and reporting of results
3	7	7. Simple random sampling
	8	8. Systematic sampling
	9	9. Stratified random sampling
4	10	10. Ratio, regression and difference estimation.
	11	11. Cluster sampling
	12	12. Two-stage cluster sampling
	13	13. Variance estimation in complex sample surveys

Module Objectives

To give you an idea of what each module is about, here is a list of the detailed objectives for each module, by topic:

Module 1

At the completion of this module you should be able to:

Topic 1: Introduction to health indicators, mortality and fertility

- Describe what health indicators are and how they are used
- Define incidence and prevalence rates, crude, central and specific rates
- Calculate age-sex-specific mortality rates
- Define fertility and infant mortality rates and how they are used as health indicators
- Construct population pyramids and use them to demonstrate the effect of fertility and mortality rates in different populations.

Topic 2: Standardisation

- Describe when to use direct and indirect standardisation
- Calculate a directly standardised rate (DSR) with appropriate confidence interval
- Calculate a directly standardised ratio with appropriate confidence interval
- Calculate an indirectly standardised ratio (ISR) with appropriate confidence interval
- Test whether an ISR is significantly different from 100
- Describe the advantages and disadvantages of each method of standardisation
- Choose the more appropriate method of standardisation for a particular problem.

Topic 3: Life tables

- Describe the difference between current and cohort life tables and the purpose of each
- Calculate a current life table, given the deaths and population figures for each age group
- Calculate the probability of dying or surviving for any number of years for people at any age
- Calculate the average number of years of life remaining to a person of any age.

Module 2

At the completion of this module you should be able to:

Topic 4: Routinely collected data and measurement of morbidity

- Describe the advantages and disadvantages of using routinely collected data, in general
- Describe the main methods of collecting morbidity data and the advantages and disadvantages of each method
- Find freely available routinely collected data about Australia's population.

Topic 5: Design of sample surveys

- Describe the target population, sampling frame, sampling unit, observation unit, selection bias and measurement bias of a given sample survey
- Design a questionnaire with well-worded questions and good layout
- Pilot test a questionnaire
- Describe some of the ethical issues you need to consider when designing and administering a questionnaire.

Topic 6: Questionnaire design, testing and reporting of results

- Describe the advantages and disadvantages of the following different modes of delivery of a survey: face-to-face, telephone, self-administered, computer-assisted administration
- Describe methods used to develop items in a questionnaire, including: using standard instruments or scales; focus groups; key informant interviews; unstructured and semi-structured interviews
- Define the various types of validity and test questionnaire items for validity and reliability
- Assess the homogeneity of a scale
- Write a survey report.

Module 3

At the completion of this module you should be able to:

Topic 7: Simple random sampling

- Describe how to take a simple random sample (SRS)
- Estimate a population mean, total and proportion and their standard errors (SE) from a simple random sample, taking into account the sampling fraction
- Estimate the required sample size for a SRS
- Compare estimates of means or proportions for independent samples and compare dependent proportions from a SRS.

Topic 8: Systematic sampling

- Describe when systematic sampling can be used instead of simple random sampling
- Estimate the population mean, total and proportion and their standard errors, for a systematic sample from a random population
- Design a repeated systematic sampling scheme for a non-random population, and estimate the population mean, total and proportion and their standard errors for the resulting sample
- Use successive differences to find a variance estimate for a systematic sample from a non-random population.

Topic 9: Stratified random sampling

- Explain the purpose of stratification and its benefits and disadvantage
- Describe how and when to carry out stratified random sampling
- Describe when and how to use proportional allocation and optimal allocation for stratified random sampling
- Estimate the population mean (or total) and the variance of the sample mean (or total) under different allocation schemes
- Estimate proportions and their standard errors from a stratified sample
- Describe post-stratification and its advantages and disadvantages, and calculate the variance of an estimate from a post-stratified sample
- Describe double (or two-phase) sampling and when it is useful, and calculate the variance of an estimate from a two-phase sample.

Module 4

At the completion of this module you should be able to:

Topic 10: Ratio, regression and difference estimation

- Describe when ratio estimation should be used
- Estimate the population ratio of two random variables, and find the standard error of this estimate, for a simple random sample or a stratified random sample
- Estimate the population mean (or total) using additional information from a correlated subsidiary variable, using ratio, regression, or difference estimation
- Estimate the sample size required to determine a population ratio, or population mean or total when ratio estimation is to be used, with the desired precision
- Calculate the relative efficiency of ratio, regression, and difference estimators.

Topic 11: Cluster sampling

- Describe how and when to carry out cluster sampling
- Explain the purpose of cluster sampling and its advantages and disadvantages
- Estimate the population mean, total or proportion and the variance of the sample mean, total or proportion for a cluster sample with clusters of equal or unequal sizes
- Estimate the population mean (or total or proportion) and its variance for a stratified cluster sample
- Design a cluster sample by determining the number of clusters of a given size required to achieve a specified bound on the error of estimation
- Design a cluster sample using sampling with probabilities proportional to the size of the cluster (pps sampling)
- Calculate the ratio estimator, unbiased estimator and pps estimator of the population total, and know when each is more appropriate.

Topic 12: Two-stage cluster sampling

- Describe how and when to carry out two-stage cluster sampling
- Calculate an unbiased estimate of the population mean (or total or proportion) and its variance for a two-stage cluster sample with clusters of equal or unequal sizes
- Calculate the ratio estimate of the population mean (or proportion) and its variance when the size of the population is unknown, and of the population total when the population size is known
- Design a two-stage cluster sample with clusters of equal size to minimise the variance of the estimate of the mean for a fixed cost, or to minimise the cost for a

fixed variance

- Design a two-stage cluster sample using first-stage sampling with probabilities proportional to the size of the cluster (pps sampling)
- Calculate an unbiased estimate of the population mean (or total or proportion) and its variance for a two-stage cluster sample in which the first stage uses pps sampling.

Topic 13: Variance estimation in complex sample surveys

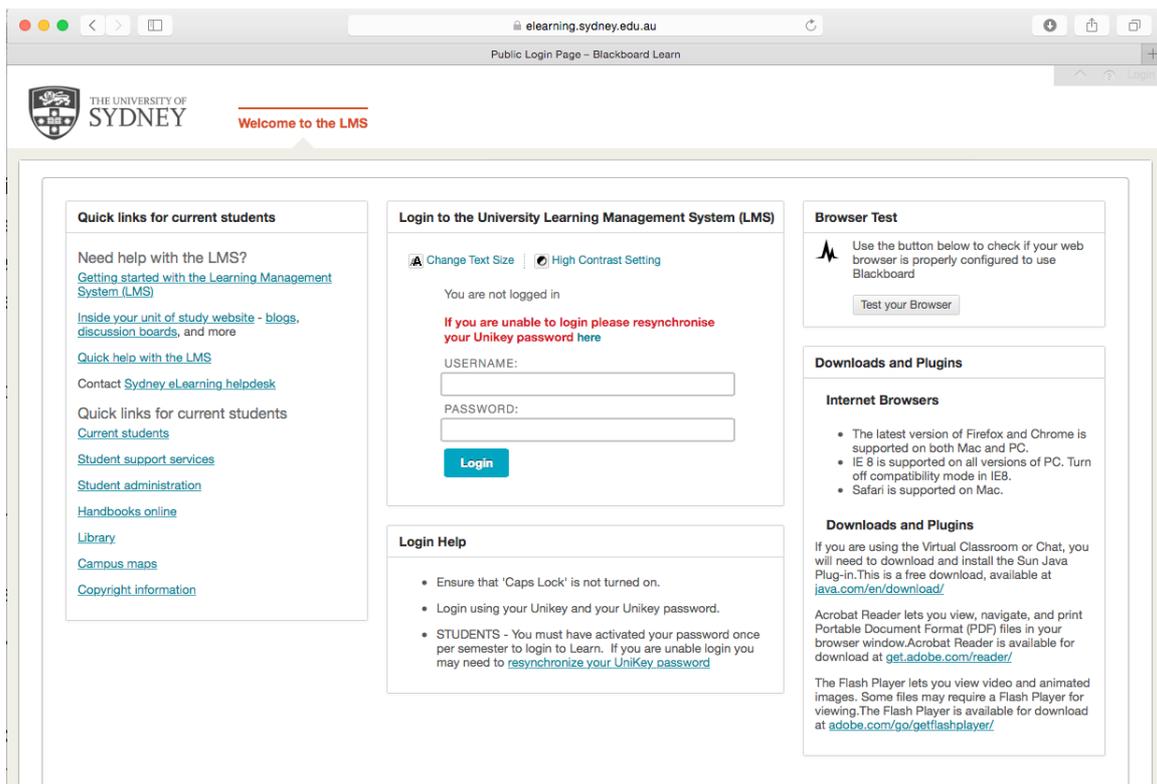
- Understand the general principle of linearization and replication
- Use the linearization method to estimate the variance of summary statistics

Methods of Communication

Online eLearning

We will use the BCA eLearning site as the main means of communication:

<http://elearning.sydney.edu.au>



Make sure that you have the **correct email registered** in the eLearning platforms or you may miss important announcements.

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.

Solutions to the weekly exercises will be posted on the HIS eLearning site. Assignments will be posted there too.

We will use the Discussions facility too. If you have a question or comment about the course material, post it to the relevant Discussion topic, where we can all access it and make a response.

- The Instructor will generally let Discussions flow between the students in each

group, except where key points seem to need resolution.

- Any general Discussion items or questions, in particular on the study guide and notes, can be posted to the other Discussion areas.

About online discussions

Discussions form an important part of your learning and your assessment. Discussions are really quite similar to face-to-face tutorials, except that your discussion is in written rather than spoken form, and you can't see those you are talking to (in fact, you may never see them). Some things to think about:

- Students will be invited by the coordinator, in a random fashion, to lead the discussion on specific exercises and share their solutions in the discussion board
- **Everybody's ideas and contributions are valuable.** We can all learn from each other's experience and insights. Don't be shy about contributing your ideas. The more ideas you contribute, the richer the discussion will be. And don't be afraid to be the first to contribute!
- **Your relationship with others in your group:** Make sure you contribute to, and read the postings in, the **Introductions** blog. Maintain good relations with the others by observing netiquette - avoiding overt criticism, flaming etc and being very careful with humour. When you can't see each other, it's easy to misunderstand something that's perhaps awkwardly worded. Learning is easier if everyone gets on well.
- **Don't be afraid to ask questions.** There may be someone else in the group wondering about just the same thing that puzzled you!
- **Interact** with the others in your group, just as you would face-to-face. Agree, disagree (politely of course, and giving reasons for your opinion). Ask for clarification, add ideas - all of this makes the discussion more interesting and worthwhile.
- **Check in often:** Get into the habit of accessing eLearning regularly and checking the 'Discussions' icon, to see if there have been any new postings. It's much easier to keep up if you check-in regularly.
- **Readings:** Be sure to do the required reading before you start the discussion, so that you can make a meaningful contribution.
- **Length of contributions:** We're not looking for assignment-length postings to discussions! We might indicate how much is needed, but if not, generally just one or two well-written paragraphs will be enough, or even one or two sentences in some cases. This is much kinder to the others in your group, who need to read what

you've written or to have a chance to make their own contribution.

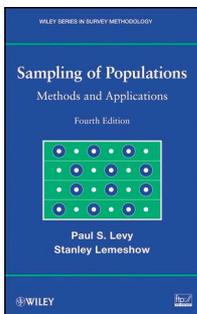
- **Getting it right:** You might like to create your posting in a word-processing program and check spelling and grammar before you post. Or type your contribution directly into eLearning and preview the message to have a look at it before you post it.
- **Not happy with your posting?** If you've posted something that you're not happy with, **you won't be able to remove it** – you'll need to ask your tutor to arrange this for you. It's better to make sure your posting is OK before you post it.
- **Adding an attachment:** You can add an attachment created in any program to your posting, but be aware that, if the people who are supposed to read the attachment don't have the same program on their computers, they won't be able to read it.

Email and Phone

If you have any questions during the semester, please email the unit coordinator, A/Prof Armando Teixeira-Pinto (armando.teixeira-pinto@sydney.edu.au) or call 02 9351 4369.

Textbook

For Module 1, no textbook is required as comprehensive notes are supplied.



For Modules 2-4, the following text is **essential, so please buy it as soon as possible:**

Levy PS and Lemeshow S. *Sampling of Populations: Methods and Applications*, Fourth Edition, Hoboken: Wiley, 2008

The price of the book varies a lot, so be sure to check several online stores. As a reference check this page:

http://www.amazon.com/gp/offer-listing/0470040076/ref=dp_olp_all_mbc?ie=UTF8&condition=all

*Disclaimer 1: I am not responsible for the information in the site above; I simply searched for the title of the book and obtained the link. Always **confirm the edition** of the book that you are buying and if it is a reliable seller!*

*Disclaimer 2: Several students complained about the quality of the **ebook** and I **do not** recommend it.*

Please note that Chapter 2 of the textbook, 'The population and the sample', is **assumed knowledge** for this unit, so you should read it to make sure you are familiar with all the material covered in this chapter.

Software

Most of the computing in the first part of this unit can easily be done using Microsoft Excel, or even on a hand calculator. **It is assumed that you know how to use Excel.** If you don't, it is strongly recommended that you take the opportunity to learn by accessing the notes available from the Student Resources part of the BCA eLearning site. There are also additional notes at the end of this section that illustrate some of the techniques in Excel that are useful for this unit.

For the analysis of surveys, it is assumed that you are **familiar with either SAS or Stata.** If not, introductory notes and purchasing details for Stata and SAS can be found in the Student Resources part of the BCA eLearning site.

Assessment

Assessment will be by four assignments, one for each module. Each assignment is worth 25% of the final mark.

Assignments will be posted online and an email will be sent to you when the assignment is posted.

Please consult the separate (pink) document 'BCA Assessment Guide' for details about submitting your assignments, and guidelines for written work.

All material submitted for assessment must be entirely your own work. Please see the note on 'Academic Dishonesty and Plagiarism' on pages 3-4 of the (pink) BCA Assessment Guide.

Assignments should preferably be submitted online. If this proves difficult then send by email as an attachment to armando.teixeira-pinto@sydney.edu.au

When you submit your assignment online, you will be required to complete a declaration, in the form of one either/or test question, certifying that you have read and understood the Academic Dishonesty and Plagiarism policy at the university in which you are enrolled. The assignment should not appear on the page until you have done this. This procedure is a compulsory requirement of all universities. See page 3 of the (pink) BCA Assessment Guide for more details.

If you don't submit your assignment via eLearning, you will need to complete the BCA Assignment Cover Sheet, a copy of which is included here after the pink pages, and fax it to us.

I strongly suggest that you keep a copy of your assignments.

It is planned that the assignments will be released and due as follows (see also Timetable) at the **end of each day**:

Assignment 1 on Module 1, released Monday 5 March, due Monday 19 March

Assignment 2 on Module 2, released Monday 19 March,

due Monday 9 April (1st part),

feedback to classmates on Monday 16 April,

and the final part on Monday 23 April (2nd part)

Assignment 3 on Module 3, released 23 Apr, due Monday 7 May

Assignment 4 on Module 4 (and 3), released Monday 21 May, due Monday 4 June.

Extensions or late submissions policy

For various reasons, you may sometimes experience difficulties in getting your assignment submitted on the due date. Requests for an extension of the due date 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 carer role. 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 granted and the new due date. Extensions will normally be no longer than three days.

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

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.

Contact details

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

Dr Armando Teixeira-Pinto, A27, University of Sydney, NSW 2006

phone 02 9351 5424 email: armando.teixeira-pinto@sydney.edu.au

fax 02 9351 5049

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

Kevin McGeechan, School of Public Health, A27, University of Sydney, NSW 2006

phone 02 9351 4648 email: kevin.mcgeechan@sydney.edu.au

fax 02 9351 5049

For **enquiries about receipt of assignments**, contact:

Biostatistics Administrative Officer, School of Public Health, University of Sydney

phone 02 9351 5994 email: sph.bsta@sydney.edu.au

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

phone 02 9562 5076 email: erica@ctc.usyd.edu.au

fax 02 9565 1863

For **enquiries about your degree program**, contact the university through which you are enrolled.