



Study Guide

Health Indicators and Health Surveys (HIS)

Semester 1, 2024

Prepared by:
Kylie-Ann Mallitt
Sydney School of Public Health, University of Sydney

Copyright © Sydney School of Public Health, University of Sydney

Contents

Contact details.....	2
Background.....	2
Context Within the Program	2
Prerequisites.....	2
Co-requisites	2
Unit summary.....	3
Workload requirements.....	3
Learning Outcomes.....	3
Unit content.....	4
Recommended approaches to study.....	4
Method of communication with coordinator	4
Module descriptions.....	4
Unit schedule	10
Assessment.....	11
Submission and academic honesty policy.....	11
Use of ChatGPT or other generative AI tools.....	12
Late submission and extension procedure	12
Learning resources	12
Software requirements and assumed knowledge	13
Required mathematical background	13
Feedback	14
Unit changes, including response to recent student evaluation	14
Acknowledgments.....	14

Health Indicators and Health Surveys (HIS)

Semester 1, 2024

Contact details

Kylie-Ann Mallitt

Sydney School of Public Health
University of Sydney

Kylie-ann.mallitt@sydney.edu.au

If you have any general BCA queries, please contact: Jaque Vaughan or Emily Higginson at the BCA Coordinating Office on 02 9562 5076/54 or via email to bca@sydney.edu.au

Background

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.

Context Within the Program

HIS is an elective unit of study within the BCA program, drawing on foundational mathematical techniques and principles developed in MFB and applying them to health-related survey and questionnaire data. Many of the data sets used across other BCA units (and in practice) will have been derived from routinely collected sources or specialised health surveys; this unit of study provides a specialised look into their development, validation, administration, and analysis considerations.

Prerequisites

None

Co-requisites

Mathematical Foundations for Biostatistics (MFB).

Unit summary

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.

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.

Learning Outcomes

At the completion of this unit students 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;
4. choose the most appropriate mode of delivery for a survey;
5. design an efficient sampling strategy to obtain a random sample of the target population;
6. estimate means, totals and proportions from survey data, taking into account the sample design, and analyse, interpret and present these results.

Unit content

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

1. Module notes describing concepts and methods, including some exercises of a more “theoretical” nature.
2. Selected readings from published articles or textbooks.
3. One or more extended examples illustrating the concepts/methods introduced in the notes and including more practically oriented exercises.

Study materials for all Modules are downloadable from the eLearning unit site. Assignments and supplementary material, such as datasets will also be posted to the unit site.

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 are encouraged to post any content-related questions to the Discussion Board, whether they relate directly to a given exercise, or are a request for clarification or further explanation of an area in the notes. The use of generative artificial intelligence is not permitted in this course.

Method of communication with coordinator

Questions about administrative aspects or course content can be emailed to the coordinator, and when doing so please use “HIS:” in the Subject line of your email to assist in keeping track of our email messages. The coordinator will be available to answer questions related to the module notes and practical exercises, and to address any other issues that require clarification.

We strongly recommend that you post content-related questions to the Discussions tool in the HIS area of BCA’s eLearning site. In 2024, we are using the Learning Management system hosted by the University of Sydney (Canvas). 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

Below is an outline of the study modules, followed by a timetable and assessment description table .

Each module is scheduled to begin on a Monday and conclude on the Sunday of the following week. **The due date for submission of the required exercises from each module is 11:59pm on the due date.**

Module 1

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

Topic 1: Introduction to health indicators and standardisation

- 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.
- 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 2: 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.

Topic 3: 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.

Module 2

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

Topic 4: 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 5: 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.

Topic 6: 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.

Module 3

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

Topic 7: 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 8: 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.

Topic 9: 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.

Module 4

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

Topic 10: 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 11: 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 12: Weighting

- Understand the principals of weighting in surveys
- Understand how weighting can reduce bias caused by deficiencies in the survey such as lack of coverage or non-response
- Calculate weights that make adjustments for the probability of selection, non-response and lack of coverage

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

Unit schedule

Semester 1, 2024 starts on Monday 26th February.

Week	Week of	Module	Topic	Assessment
1	26 February	Module 1	Introduction to health indicators and standardisation	
2	4 March	Module 1	Life tables	Assignment 1 released 4 March
3	11 March	Module 1	Routinely collected data and measurement of morbidity	
4	18 March	Module 2	Design of sample surveys	Assignment 1 due 18 March.
5	25 March	Module 2	Questionnaire administration, testing and reporting of results	Assignment 2 released 25 March
	1 April		mid semester break 1 week only	
6	8 April	Module 2	Simple random sampling	Assignment 2 (Part 1A) due 8 April
7	15 April	Module 3	Systematic sampling	Feedback provided for Assignment 2 (Part 1A) Monday 15 April
8	22 April	Module 3	Stratified random sampling	
9	29 April	Module 3	Ratio and regression estimation	Assignment 3 released 29 April. Assignment 2 (remainder) due 29 April.
10	6 May	Module 4	Cluster sampling	
11	13 May	Module 4	Two-stage cluster sampling	Assignment 3 due 13 May
12	20 May	Module 4	Weighting	Assignment 4 released 20 May
13	27 May	Module 4	Variance estimation in complex sample surveys	
	3 June			Assignment 4 due 3 June

Assessment

Assessment will include 4 written assignments worth 25% each, to be made available in the middle and at the end of the semester, and to be completed within approximately 2 weeks. These assignments will be posted on the eLearning site together with an online Announcement broadcasting their availability.

Assessment name	Assessment type	Coverage	Learning objectives	Weight
Major Assignment 1	Assignment	Modules 1	1,2	25%
Major Assignment 2	Assignment	Module 2	3,4	25%
Major Assignment 3	Assignment	Module 3	5,6	25%
Major Assignment 4	Assignment	Modules 4	5,6	25%

In general you are required to submit your work typed in Word or similar (e.g. using Microsoft's Equation Editor for algebraic work) and we strongly recommend that you become familiar with equation typesetting software such as this. If extensive algebraic work is involved you may submit neatly handwritten work, however please note that marks will potentially be lost if the solution cannot be understood by the markers due to unclear or illegible writing. This handwritten work should be scanned and collated into a single pdf file and submitted via the eLearning site. 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 and academic honesty policy

All assessment material should be submitted via the relevant Assessment module in Canvas unless otherwise advised. Turnitin plagiarism detection is applied to all submissions. For detailed information, please see the [BCA Assessment Guide](#), which includes links to the Academic Honesty policies at member universities. Please familiarise yourself with the procedures and policies at your home university. You will need to indicate your compliance with the plagiarism guidelines and policy at your university of enrolment before making the submission.

A special note regarding "contract cheating" sites: Unfortunately there have been instances in the past of students using such websites to post assignment questions and receive solutions (usually for a fee). We have arrangements with these sites to identify the students posting questions or accessing the solutions, and such students will be referred to and face disciplinary processes at their home university.

Use of ChatGPT or other generative AI tools

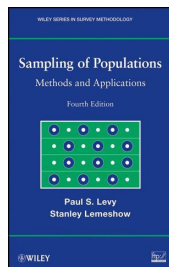
The assessment tasks in this Unit have been designed to be challenging, authentic and complex. The use of generative artificial intelligence is prohibited for all assessments in this course.

Late submission 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

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:

- https://www.bookdepository.com/Sampling-Populations-Paul-S-Levy/9780470040072?ref=pd_detail_1_sims_b_p2p_1

*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.*

Note, however, it is available as an ebook through each of the BCA University libraries. Links

University of Sydney

https://sydney.primo.exlibrisgroup.com/discovery/fulldisplay?docid=alma991027629159705106&context=L&vid=61USYD_INST:sydney&search_scope=MyInst_and_CI&isFrbr=true&tab=Everything&lang=en

University of Adelaide

https://adelaide.hosted.exlibrisgroup.com/primo-explore/fulldisplay?docid=ALMA51183368310001811&context=L&vid=SUA_NEW&search_scope=Everything&tab=default_tab&lang=en_US

Monash University

https://monash.hosted.exlibrisgroup.com/primo-explore/fulldisplay?docid=catau51307305830001751&context=L&vid=MONUI&search_scope=au_everything&isFrbr=true&tab=default_tab&lang=en_US

Macquarie University

https://multisearch.mq.edu.au/permalink/61MACQUARIE_INST/7h5qs5/alma99241216710802171

University of Queensland

https://search.library.uq.edu.au/primo-explore/fulldisplay?docid=61UQ_ALMA51162061570003131&context=L&vid=61UQ&search_scope=61UQ_All&isFrbr=true&tab=61uq_all&lang=en_US

Note the textbook is only available as a hardcopy at the University of Melbourne library.

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 requirements and assumed knowledge

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.

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

For help with R, please see [Learning R](#) in the Student Resources site.

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.

Required mathematical background

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>

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
- Feedback from non-assessed online quizzes
- Responses to questions posted on the Discussion Board

Your feedback to us:

One of the formal ways students provide feedback on teaching and their learning experience is through the BCA student evaluation survey at the end of each semester. The feedback is anonymous and provides the BCA with evidence of aspects that students are satisfied with and areas for improvement.

Unit changes, including response to recent student evaluation

HIS was last delivered in Semester 1 2023. In response to student feedback in 2022, Assignment 2 has been modified to ensure equity.

Acknowledgments

These notes include material that was previously developed by Professors Judy Simpson and Richard Taylor, Sydney School of Public Health, University of Sydney.