



Centre for Big Data Research in Health

Professional Development Courses in Health Data Science

As one of the world's top 50 universities, UNSW Sydney is globally recognised for innovative teaching and world-leading research. We house Australia's only Centre for Big Data Research in Health, which aims to maximise the use of health data at-scale to transform disease prevention and management and deliver the highest quality health services for people globally. As well as being an international hub for cross-disciplinary research using big data, the Centre for Big Data Research in Health trains the future health workforce of health data scientists through its Australian-first postgraduate programs and suite of professional development courses.



UNSW
SYDNEY

Australia's
Global
University



CENTRE FOR
BIG DATA RESEARCH
IN HEALTH

Healthcare needs Health Data Scientists

THE DATA DELUGE



33 TRILLION GB

The total amount of data produced in human history up to 2018.



172.2 MILLION

The number of wearable fitness trackers sold worldwide in 2018



75 DAYS

How long it takes for global health data to double in size



175 TRILLION GB

The total amount of data humans will have produced by 2025. If that much info was contained in a string of USB each storing 128GB, it would reach the moon and back almost 100 times.

NEED FOR BIG DATA ANALYTICS IN HEALTH



2.5 X

The rate Indigenous Australians are admitted to hospitals compared to non-Indigenous Australians - a problem Health Data Scientists at UNSW are working to address



204 MILLION

The number of prescriptions issued in Australia in 2017-2018



11.3 MILLION

The number of recorded hospitalisations in Australia in 2016-2017



1 YEAR

How long it can take for infectious disease researchers to gain access to death certificates, which seriously hinders our ability to respond to new epidemics

HEALTH DATA SCIENTISTS IN DEMAND



#1

Where career experts Glassdoor.com ranked data scientist in their US job rankings for 2016, 2017, 2018 and 2019, based on job satisfaction, number of job openings and salary



Top 25

LinkedIn's Most In-Demand Hard Skills 2019 ranking of Analytical Reasoning, Scientific Computing and Data Science



1.7 MILLION

The estimated shortfall in the number of data analysts required in the US in 2018



\$130,000

The median annual salary for analytics professionals in Australia

Professional Development in Health Data Science

Effective learning solutions to meet health analytics competency needs

In today's world where there is growing demand for data-driven solutions, our Health Data Science professional development courses provide a timely option for employers seeking to quickly upskill staff and for individuals wanting to gain new skills to meet immediate work demands or support future career aspirations. The Health Data Science professional development courses stem from our Master of Science in Health Data Science program at UNSW Sydney.

Our courses aim to build and equip graduates with essential competencies for which there is high demand in the health data scientist workforce. Teaching examples are all health-specific using relevant real-world Australian health system data where possible. Curriculum has been developed in partnership with cross-disciplinary UNSW experts from the Centre for Big Data Research in Health, School of Computer Science and Engineering, School of Mathematics and Statistics, Ingham Institute for Applied Medical Research and key staff of the Australian Government Department of Health.

Benefits of our professional development program

- A learning journey for understanding complex health issues and the generation of innovative practical solutions
- Competency focussed, including skills in R, Python, SAS, data wrangling and management, statistical modelling, machine learning and data visualisation
- Cutting-edge and industry standard software and programming practices
- Interactive and engaging online learning experience with learning partner OpenLearning (www.openlearning.com)
- Expert faculty with extensive experience of Australian health data and advanced health analytics

Health Big Data

Big data have no agreed definition, but the term is in general applied to data that by virtue of their size and/or complexity pose challenges to traditional methods for management and analysis. In health, such data include the millions of records that are generated routinely by health services, real-time clinical data captured at the point-of-care, genomic data produced in research and clinical settings, and health-related data generated by the population at large through technologies such as wearable devices and social media. How we make use of this data and transform it into action to better support clinical care, inform health policy and improve population health has never been more important than now.

What is Health Data Science?

Health Data Science is the science and art of generating data-driven solutions through comprehension of complex real-world health problems, employing critical thinking and analytics to derive knowledge from health big data. Health Data Science is an emerging exciting discipline that spans the domain areas of health and medicine, statistics and computer science.

Health Data Science Courses

Teaching Approach

Each course is organised into ten chapters with a schedule that is aligned to the UNSW Term calendar. Content is delivered fully online using a combination of instructional videos, readings and interactive exercises that aims to build analytics skills, stimulate critical thinking and engage peer to peer learning. The courses use a variety of assessment modalities including multiple choice question gamification, data management plans, algorithm challenges, design portfolios and reflective blogs.

The notional study time commitment for a course session is 10 hours per week.

Assumed Knowledge

Statistical Modelling 1 & 2, and Machine Learning and Data Mining are Graduate Diploma level courses that build upon the Graduate Certificate level courses of Statistical Foundations for Health Data Science and Computing for Health Data Science. Future students are encouraged to contact cbdrh@unsw.edu.au to confirm sufficiency of knowledge base before enrolling in these courses, which require programming skills in R and/or Python.

Visualisation and Communication of Health Data requires R programming skills. Future students can contact cbdrh@unsw.edu.au to learn how proficiency can be gained through pre-course R learning modules.

Context of Health Data Science

Graduate Certificate level course

Statistical Foundations for Health Data Science

Graduate Certificate level course

Management and Curation of Health Data

Graduate Certificate level course

Computing for Health Data Science

Graduate Certificate level course

Machine Learning and Data Mining

Graduate Diploma level course

Statistical Modelling I

Graduate Diploma level course

Visualisation and Communication of Health Data

Graduate Diploma level course

Statistical Modelling II

Graduate Diploma level course

Health Data Science Courses

Context of Health Data Science

Graduate Certificate level course

The Context of Health Data Science provides an introduction to how data are generated and used in a contemporary health system. We look at how health outcomes can be measured and reported in various forms of health data, and how these health data can reveal inequalities in health. The course describes the major sources of health data, including those relating to primary care, hospital stays and prescription medicines, and how this (and other) information can be used by the health data scientist to create evidence for policy and research.

Activities are structured to foster a scientific, questioning attitude in the student. Students are encouraged to think critically about how health data are recorded, what this reveals about the underlying health delivery systems, and be creative in their use of health data sources to create or critically appraise evidence.

Assumed Knowledge

None

Assessment

- 9 End of chapter activities

Session

Term 1: 17 February to 16 May 2020

Statistical Foundations for Health Data Science

Graduate Certificate level course

Almost all aspects of health data science, from the most basic descriptive analyses through to the development of the most sophisticated deep learning models, are built on a set of foundation statistical concepts and principles, encompassing both frequentist and Bayesian paradigms.

The course will provide the student with a thorough understanding of the Law of Large Numbers and the Central Limit Theorem, probability distributions, likelihood and likelihood estimation, Bayes theorem and Bayesian estimation, Monte Carlo methods and resampling methods such as the bootstrap, frequentist inference, and essential epidemiological and study design concepts. The approach is highly computational. Rather than relying on mathematical proofs and theorems, students investigate and verify these concepts through simulations which they construct themselves, while simultaneously gaining proficiency in the widely-used, open-source R statistical programming language. The end result is a sound knowledge of statistical computing and good programming practice, allied to a hands-on understanding of the statistical underpinnings of both regression modelling and machine learning.

Assumed Knowledge

None

Assessment

- 10 End of chapter activities
- 1 Course reflection

Session

Term 1: 17 February to 16 May 2020
 Term 3: 14 September to 12 December 2020

Health Data Science Courses

Management and Curation of Health Data

Graduate Certificate level course

This course is designed to equip students with the skills required to collect or obtain data, design data management strategies aligned with best practice, and appreciate the day to day practicalities of data curation for sound data management. Students will develop data wrangling skills required to assemble data suitable for analysis and research purposes. Data wrangling skills will focus on the key areas of data security, data exploration, documentation of data (for example data dictionaries) and data management, with the ultimate aim of creating analysis-ready datasets and ensuring reproducible results.

The course is taught using SAS.

Assumed Knowledge

None

Assessment

- 2 Assignments

Session

Term 2: 1 June to 29 August 2020
 Term 3: 14 September to 12 December 2020

Computing for Health Data Science

Graduate Certificate level course

Computing is an integral part of current day health services delivery and management. Its importance continues at an ever increasing pace. Amongst other areas, computing is used in data collection, monitoring, analysis and visualisation. The ability to use computers to help solve problems efficiently, is therefore a very important (critical) skill for a Health Data Scientist. The aim of this course is to introduce computing for Health Data Science. An emphasis is placed on practical computational problem solving. To realise this aim, students will learn to program in the Python language by tackling health-related problems. Solving these problems will require use of some Python packages (e.g. NumPy). Specifically, this course will equip you with the skills to design and implement algorithms to solve computational problems posed within contextual health examples. Theoretical design principles are reinforced with extensive ‘hands-on’ coding in Python.

Assumed Knowledge

None

Assessment

- 9 Exercises
- 1 Mid-term assignment
- 1 Project assignment

Session

Term 1: 17 February to 16 May 2020

Health Data Science Courses

Machine Learning and Data Mining

Graduate Diploma level course

Machine learning and data mining bring together methods coming from statistics and computer science and apply these to databases both large and small. They use a powerful and diverse set of techniques and algorithms to discover patterns and relationships in data with the final goal of creating knowledge from these data. These methods are increasingly being applied to the vast amounts of health data that are generated through sources including electronic medical records, medical and pharmaceutical claims, medical imaging, wearable and implantable devices and social media.

This course provides an introduction to data mining and machine learning, including both supervised and unsupervised techniques. You will learn about the underlying theory, as well as gain the practical know-how required to effectively apply these techniques to real-world health datasets to answer new health data science problems. The widely-used, open-source Python programming language is used to teach the course.

Assumed Knowledge

- Statistical Foundations for Health Data Science or equivalent
- Computing for Health Data Science or equivalent
- Python coding proficiency
Contact cbdrh@unsw.edu.au about how to gain Python proficiency

Assessment

- 8 Exercises
- 1 Project report

Session

Term 2: 1 June to 29 August 2020
Term 3: 14 September to 12 December 2020

Statistical Modelling I

Graduate Diploma level course

This course provides a sound grounding in the theory and practice of fitting statistical regression models, with particular focus on the flexibility of generalised linear models (GLMs). Starting with linear regression, a major theme of the course is best practice in model fitting, including thorough exploratory data analysis, model assumption checking, data preparation and transformation, including the use of imputation, and careful attention to model adequacy and diagnostics. Emphasis is given to content-aware, purposive model building and the use of Directed Acyclic Graphs (DAGs) of causal relations to inform model parameter selection. Non-linear, logistic, binomial and Poisson models for count data are also covered. Effect modifications (interactions) and their meaning in a health context are explored. The presentation and visualisation of statistical models is considered, with emphasis on the explanatory insights that can be gained from well-constructed models. The final part of the course covers basic time-series models, survival analysis and other time-to-event models. The course is taught using the R programming language.

Assumed Knowledge

- Statistical Foundations for Health Data Science or equivalent
- R coding proficiency
Contact cbdrh@unsw.edu.au about how to gain R proficiency

Assessment

- 3 Exercises
- 1 Project report

Session

Term 2: 1 June to 29 August 2020

Health Data Science Courses

Visualisation and Communication of Health Data

Graduate Diploma level course

Health Data Scientists present information to audiences across a range of backgrounds, spanning a spectrum from naïve or non-practitioners of a discipline to highly informed and expert audiences. Effective communication across different media types is essential. Appropriate data visualisation techniques can greatly increase the effectiveness of communication. An understanding of some basic simple techniques can ensure communication remains effective across diverse audiences. An understanding of the computation and presentation aspects of health data visualisation can increase not only the effectiveness of communication but also the efficiency of work effort.

This course takes a practical approach to creating appropriate, reproducible and transparent analyses and visualisations. Using R and RStudio, it develops useful data science analysis and visualisation techniques for different types of data visualisation and communication, including charts and graphs and written and oral communication forms. What makes a good map is discussed and the use of geospatial information is explored through the construction of an interactive Shiny application.

Assumed Knowledge

- R coding proficiency
- Contact cbdrh@unsw.edu.au about how to gain R proficiency.

Assessment

- 7 Exercises
- 1 Group project

Session

Term 2: 1 June to 29 August 2020

Statistical Modelling II

Graduate Diploma level course

Sophisticated modelling techniques are essential for the analysis of real-world health data. Building on Statistical Modelling I, this course expands the statistical toolkit and broadens students' understanding of relevant statistical approaches for the analysis of realistically complex data structures and research questions. The course is aimed at those currently working or planning on working in health or a health-related field, and who are interested in applying advanced statistical methods to analyse complex data.

Topics covered in this course include multilevel models for hierarchical data; analysis of time series and longitudinal data; quasi-experimental approaches for drawing causal inferences from observational data; multiple imputation for missing values; and simulation approaches for study planning and model evaluation.

Content is delivered through a combination of online readings, expert lectures and interactive tutorials. Statistical concepts are illustrated with a variety of health examples, and students will learn how to implement methods using leading statistical software.

Assumed Knowledge

- Statistical Foundations for Health Data Science or equivalent
 - Statistical Modelling I or equivalent
 - R coding proficiency
- Contact cbdrh@unsw.edu.au about how to gain R proficiency

Assessment

- 3 Exercises
- 1 Project report

Session

Term 3: 14 September to 12 December 2020

Course Enrolment and Fees

To enrol into courses for 2020, click on the course name to be directed to the registration and payment site.

Term 1 - 17 February to 16 May 2020

- Context of Health Data Science
 - Statistical Foundations for Health Data Science
 - Computing for Health Data Science
-

Term 2 - 1 June to 29 August 2020

- Management and Curation of Health Data
 - Statistical Modelling 1
 - Visualisation and Communication of Health Data
 - Machine Learning and Data Mining
-

Term 3 - 14 September to 12 December 2020

- Machine Learning and Data Mining
- Statistical Modelling 2
- Statistical Foundations for Health Data Science
- Management and Curation of Health Data

Course enrolment closes one week before the course start date.

There are no entry requirements to enrol into Health Data Science professional development courses, enrolment is open to Australian and international professionals and students.

Check you meet the assumed knowledge requirements for courses before enrolling. If in doubt, email cbdrh@unsw.edu.au to confirm.

Following course enrolment and payment, a notification of registration and receipt will be emailed from UNSW. A few days prior to the start of the course, an invitation email from OpenLearning will be sent inviting you to access the course at www.openlearning.com.

Course Enrolment and Fees

Each course costs AU \$3400 (including GST). The cost includes access to all course content online, online facilitation by UNSW faculty and course tutors, assessment marking and feedback, and dependent on successful completion, course certificates.

No refunds can be issued for course withdrawals. However, requests for transfers to an alternative Health Data Science professional development course can be considered if submitted within the period of chapters 1 and 2 of the enrolled course. Send transfer requests to cbdrh@unsw.edu.au.

Transferring into UNSW Health Data Science postgraduate programs

Each of the Health Data Science professional development courses are equivalent to 6 Units of Credit (UoC) of a UNSW postgraduate course in Health Data Science and they can be used as recognition of prior learning towards a UNSW postgraduate qualification in Health Data Science.

At the time of admission to UNSW, students can apply to have Health Data Science professional development courses that they have completed (must have scored at least 50%) recognised as advanced standing or credit transferred to the UNSW degree program. Up to 50% of the total UoC of the program can be transferred and students must then complete at least 50% of the remaining UoC as a UNSW enrolled student to be awarded a UNSW qualification. For example, for the Graduate Certificate (24 UoC, 4 courses) you can use up to two Graduate Certificate level courses from the professional development courses (12 UoC) and then complete 2 courses (12 UoC) as a UNSW enrolled student.

Learn more about the UNSW postgraduate programs in Health Data Science at cbdrh.med.unsw.edu.au/postgraduate-coursework

For more information about Professional Development courses in Health Data Science, please contact us:

Phone: +61 2 9385 9064

Email: cbdrh@unsw.edu.au

Web: cbdrh.med.unsw.edu.au

Visit: Level 4, Centre for Big Data Research in Health, Lowy Building C25, UNSW Sydney, Kensington, Australia

