

GRADUATE DIPLOMA
MODULE 5 - TOPICS IN APPLIED STATISTICS

INTRODUCTORY NOTE 1: This module covers four application areas: 1) multivariate methods, 2) censored data: survival & reliability, 3) demography & epidemiology and 4) sampling. This reading list consists of books, some available as eBooks, as well as internet resources. Internet resources consist of books (many available for free download), online courses, and online articles or short guides on particular topics.

INTRODUCTORY NOTE 2: Disclaimer: While the RSS has made every effort to ensure that the material and information in this document is accurate and up-to-date when published, it is only general information and may be out-of-date when accessed. The websites referenced in the document may provide the user with the ability to download materials. The rules and permissions for downloading materials are site specific and the user's sole responsibility, not that of the Royal Statistical Society.

BOOKS

The books are listed in alphabetical order by name of first author. Some are general texts that cover much or all of the syllabus for the module; others are focused on particular areas. Some of books are available as eBooks at a reduced price. You may also be able to purchase individual chapters and/or purchase online access for a limited period, e.g. 24 hours.

Practical Statistics for Medical Research <http://www.crcpress.com/product/isbn/9781439856789> Altman DG (1990) Practical Statistics for Medical Research. Chapman & Hall/CRC Texts in Statistical Science, 624 pages. Also available as eBook.

Statistical Methods in Medical Research, Fourth Edition <http://onlinelibrary.wiley.com/book/10.1002/9780470773666>

Armitage P, Matthews JNS and Berry G (2008) Statistical Methods in Medical Research, Fourth Edition. Wiley-Blackwell, 832 pages. Also available as eBook.

Sample Survey Principles and Methods, Third Edition <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470685905.html>

Barnett V (2002) Sample Survey Principles and Methods, Third Edition. Hodder Arnold, 258 pages

Introduction to Multivariate Analysis
<http://www.crcpress.com/product/isbn/9780412160400>

Chatfield C and Collins AJ (1981) Introduction to Multivariate Analysis. Chapman & Hall/CRC Texts in Statistical Science, 248 pages.

Sampling Techniques
<http://www.crcpress.com/product/isbn/9780412160400>
Cochran WG (1977) Sampling Techniques, Third Edition. Wiley, 448 pages.

Modelling Survival Data in Medical Research, Third Edition
<http://www.crcpress.com/product/isbn/9781439856789>
Collett D (2014) Modelling Survival Data in Medical Research, Third Edition. Chapman & Hall/CRC Texts in Statistical Science, 548 pages. Also available as eBook.

Analysis of Survival Data
<http://www.crcpress.com/product/isbn/9780412244902>
Cox DR and Oakes D (1984) Analysis of Survival Data. Chapman & Hall/CRC Monographs on Statistics & Applied Probability, 212 pages.

Statistical Analysis of Reliability Data

<http://www.crcpress.com/product/isbn/9780412594809>

Crowder MJ, Kimber AC, Smith RL and Sweeting TJ (1994) Analysis of Survival Data. Chapman & Hall/CRC Texts in Statistical Science, 264 pages.

Demographic Methods <http://www.sponpress.com/books/details/9780340718926/>

Hinde A (1998) Demographic Methods. Routledge, 320 pages. Also available as eBook.

Multivariate Statistical Methods: A Primer, Third Edition <http://www.crcpress.com/product/isbn/9780412594809>

Manly BFJ (2004) Multivariate Statistical Methods: A Primer, Third Edition. Chapman & Hall/CRC Textbook, 224 pages. Also available as eBook.

Survival Analysis: A Practical Approach, Second Edition <http://onlinelibrary.wiley.com/book/10.1002/0470034572>

Machin D, Cheung YB, and Parmar MKB(2004) Survival Analysis: A Practical Approach, Second Edition. Wiley. Also available as eBook.

Demographic Techniques, Fourth Edition <http://www.crcpress.com/product/isbn/9781439839706>

Pollard AH, Yusuf F and Pollard GM (1990) Demographic Techniques, Third Edition. A.C. Wilson, Sydney [previous editions by Pergamon Press Australia] 185 pages.

Demography: Measuring and Modeling Population Processes <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-1557864519.html>

Preston SH, Heuveline P, Guillot M (2009) Demography: Measuring and Modeling Population Processes. Wiley-Blackwell Publishing. 308 pages.

Epidemiology: Study Design and Data Analysis, Third Edition <http://www.crcpress.com/product/isbn/9781439839706>

Woodward M (2013) Epidemiology: Study Design and Data Analysis, Third Edition. Chapman & Hall/ CRC Texts in Statistical Science, 898 pages. Also available as eBook.

ONLINE COURSES ON DEMOGRAPHY AND EPIDEMIOLOGY

Introduction to Demographic Methods <http://ocw.jhsph.edu/index.cfm/go/viewCourse/course/demographicmethods/coursePage/index/>

Lecture slides and accompanying audio are presented for each session along with assignments and solutions.

- Lecture 6: Standardisation
- Lecture 7: Life Table

Population Analysis for Policies & Programmes

<http://papp.iussp.org>

This course introduces the methods used to analyse population data, the sources of this data and the methods used to collect this data. There are exercises with answers. It includes a glossary, and modules:

- PAPP101: Introduction to Demographic Analysis, including
 - Session 02: How to measure demographic events - covers incidence versus prevalence
 - Session 03: How demographers think about populations: age and sex - covers population pyramids – Session 06: Measurement of mortality and standardisation - covers standardised mortality ratio – Sessions 07 & 08: Life tables I & II - covers the construction and use of life tables.

- PAPP102: Demographic Data: Sources, Collection and Evaluation,
 - Session 01: Censuses and vital registration
 - Session 02: Single-round surveys
 - Session 03: Longitudinal studies
 - Session 04: Design and conduct of surveys
 - Session 05: Conducting quantitative demographic inquiries: sampling
- PAPP103: Demographic Analysis: Further Methods and Models, and
- PAPP104: Demographic Analysis: Applications and Extensions.
 - Session 02: Life tables from survival data

STAT 507 Epidemiological Research Methods <https://onlinecourses.science.psu.edu/stat507/>

A basic introductory course on epidemiological research methods in 13 lessons. While some statistical models are discussed, they are not discussed in detail. For those who know the R statistics software, a Little Book of R for Biomedical Statistics by Avril Coghlan, would complement this course material. <https://a-little-book-of-r-for-biomedical-statistics.readthedocs.org/en/latest/>

Fundamentals of Epidemiology I & II <http://ocw.jhsph.edu/index.cfm/go/viewCourse/course/FundEpi/coursePage/index/http://ocw.jhsph.edu/index.cfm/go/viewCourse/course/fundepiii/coursePage/index/>

Supercourse on Epidemiology

<http://www.pitt.edu/~super1/assist/topicsearch.htm>

Repository of lectures on global health and prevention. It is a network of over 56000 scientists in 174 countries who are sharing for free a library of 5805 lectures in 33 languages. Search for lectures using: the Keyword Index. See Topic Index especially sections on Biostatistics, Epidemiology, Research Methods, and Statistics.

ONLINE BOOKS ON EPIDEMIOLOGY

A Short Introduction to Epidemiology, Second Edition <http://csm.lshtm.ac.uk/files/2010/09/A-Short-Introduction-to-Epidemiology-Second-Edition.pdf> Pearce N (2005) Published as an Occasional Report Series No 2, Centre for Public Health Research, Massey University, Wellington, New Zealand, pp 130. ISBN 0-476-01236-8.

Understanding the Fundamentals of Epidemiology – An Evolving Text <http://www.epidemiolog.net/evolving/TableOfContents.htm>

Schoenbach VJ with Rosamond WD (Fall 2000 Edition) pp 572. This online text includes assignments and solutions. It includes the standardisation of rates and ratios as well as the standardised mortality ratio. While some statistical models are discussed, they are not discussed in detail.

The International Agency for Research on Cancer (IARC) has a number of publications on epidemiology and statistical methods in cancer research, including survival analysis, which may be downloaded for personal research purposes. Some of the statistical material is advanced, but the basics of the techniques discussed are covered in depth.

Cancer Epidemiology: Principles and Methods <http://www.iarc.fr/en/publications/pdfs-online/epi/cancerepi/index.php>

Dos Santos Silva I (Editor) (1999) International Agency for Research on Cancer, Lyon. France, pp 441. Chapter 12 is an introduction to survival analysis.

Statistical Methods in Cancer Research Volume I - The Analysis of Case-control Studies

<http://www.iarc.fr/en/publications/pdfs-online/stat/sp32/index.php>

Breslow NE & Day NE (Editors) (1980) IARC Scientific Publications No. 32, pp 346.

Statistical Methods in Cancer Research Volume II - The Design and Analysis of Cohort Studies [http://](http://www.iarc.fr/en/publications/pdfs-online/stat/sp82/index.php)

www.iarc.fr/en/publications/pdfs-online/stat/sp82/index.php

Breslow NE & Day NE (Editors) (1987) IARC Scientific Publications No. 82, pp 415.

Statistical Methods in Cancer Research Volume IV - Descriptive Epidemiology <http://www.iarc.fr/en/publications/pdfs-online/stat/sp128/index.php>

Esteve J, Benhamou E & Raymond L (Editors) (1994) IARC Scientific Publications No. 128, pp 313.

Chapter 2 includes a section on standardised rates. Chapter 4 Techniques for survival analysis includes estimation of the survival distribution and methods of comparison such as the Cox model.

ONLINE COURSES ON STATISTICS

Statistics and demography courses with open online course materials have been produced by the Johns Hopkins Bloomberg School of Public Health (JHSPH). Lecture slides and accompanying audio are presented for many of the courses. Most courses have practice problems with solutions. Browse the open courses at <http://ocw.jhsph.edu/index.cfm/go/find.browse#courses>

Introduction to Biostatistics <http://ocw.jhsph.edu/index.cfm/go/viewCourse/course/introbiostats/coursePage/index/>

Methods in Biostatistics I & II <http://ocw.jhsph.edu/index.cfm/go/viewCourse/course/MethodsInBiostatisticsI/coursePage/index/><http://ocw.jhsph.edu/index.cfm/go/viewCourse/course/methodsinbiostatisticsii/coursePage/index/>

STAT 507 Analysis of Discrete Data <https://onlinecourses.science.psu.edu/stat507/>

- Lesson 3: Two-way tables. Covers estimation and confidence intervals for 2 by 2 tables.
- Lesson 5: Three-way tables. Covers the Cochran-Mantel-Haenszel test procedure.

ONLINE COURSES ON SAMPLING

Issues in Survey Research Design <http://ocw.jhsph.edu/index.cfm/go/viewCourse/course/surveyresearchdesign/coursePage/index/>

- Lecture 1: Overview of Course and the Survey Design Process
- Lecture 3: Selecting Target Population and Sampling Frame

Statistical Methods for Sample Surveys <http://ocw.jhsph.edu/index.cfm/go/viewCourse/course/StatMethodsForSampleSurveys/coursePage/index/>

- 7 lecture course - covers simple random, systematic, stratified and cluster sampling.
- Also covers weighting and survey errors.

Rapid Epidemiological Surveys in Developing Countries <http://www.ph.ucla.edu/epi/rapidsurveys/RSCourse/>

- 17 lecture course - covers simple random, systematic, stratified and simple cluster sampling, but the focus is on two-stage cluster surveys.

“The intention of the course was to present the need for rapid surveys; an appreciation of theories of sampling; the nature of rapid two-stage cluster surveys; and the importance of such surveys for decision-making.”

Sample Survey Design and Analysis: A Comprehensive Three-Day Course, with Application to Monitoring and Evaluation <http://www.foundationwebsite.org>

- Day 1: Basic Concepts of Sample Survey, which includes introduction and syllabus
- Day 2: How to Design Surveys and Analyze Survey Data
- Day 3: Special Topics Practical Problems in Survey Design

ONLINE BOOKS & SHORT GUIDES ON SAMPLING

Survey Methods and Practices <http://www.statcan.gc.ca/pub/12-587-x/12-587-x2003001-eng.pdf>

“This publication shows readers how to design and conduct a census or sample survey. It explains basic survey concepts and provides information on how to create efficient and high quality surveys. It is aimed at those involved in planning, conducting or managing a survey and at students of survey design courses.”
Statistics Canada (2010) Survey Methods and Practices. Statistics Canada, Catalogue no. 12-587-X, Ottawa.

A Guide to Good Survey Design (3rd ed)

<http://www.stats.govt.nz/methods/survey-design-data-collection.aspx>

“This guide is for those who undertake or commission surveys. It identifies the issues associated with planning, undertaking, commissioning, managing, and processing a survey.”
Statistics NZ (2014).

Sampling for Surveys - A Short Guide

<http://msor.rsscse.org.uk/leaflets/>

“In this short guide, we have discussed many basic concepts in finite population sampling: considering the defining issues of inference for finite populations, the distinctive fundamental sampling schemes and many of the practical results for taking samples and inferring properties of the population.”

Barnett V, Maths, Stats & OR Network. For more detailed study, see his book. Barnett V (2002) Sample Survey Principles and Methods 3rd Ed. Arnold, London

Sampling Methods and Planning a Survey

http://courses.ncssm.edu/math/Stat_Inst/Stats2007/2007_statistics_institute.htm

No authors listed, from a talk by Richard L., Scheaffer, North Carolina School of Science and Mathematics, 2007 Statistics Institute, 9 pages.

Sampling Methods and Practice http://courses.ncssm.edu/math/Stat_Inst/Notes.htm

"This 36 page paper gives a overview and comparison of the different standard methods of conducting probability samples, the steps in planning a survey and estimation methods as well as the problems of sample size and allocation." Scheaffer RL (2007) North Carolina School of Science and Mathematics Statistics Leadership Institute Notes, July 1999. 36 pages.

Estimating Parameters (Simple, Stratified, and Cluster Samples) http://courses.ncssm.edu/math/Stat_Inst/Stats2007/2007_statistics_institute.htm

"This paper gives the computational formulas for estimating means and proportions from simple, stratified, and cluster samples. Some formulae are derived, others given. Ratio estimation is also included." No authors listed, North Carolina School of Science and Mathematics, 2007 Statistics Institute, 19 pages.

ONLINE ARTICLES ON DEMOGRAPHY & EPIDEMIOLOGY

British Medical Journal: Statistics Notes

<http://www-users.york.ac.uk/~mb55/pubs/pbstnote.htm> <http://www.bmj.com/specialties/statistics-notes>

"Series of short (many one-page!) articles on statistics"

- Diagnostic tests 1: sensitivity and specificity – Altman DG & Bland JM BMJ (1994) 308, 1552
- Diagnostic tests 2: predictive values - Altman DG & Bland JM BMJ (1994) 309, 102
- Diagnostic tests 3: receiver operating characteristic plots - Altman DG & Bland JM BMJ (1994) 309, 188
- Time to event (survival) data - Altman DG & Bland JM BMJ (1998) 317: 468-469
- Survival probabilities (the Kaplan-Meier method) - Bland JM & Altman DG BMJ (1998) 317, 1572-1580
- The logrank test - Bland JM & Altman DG BMJ (2004) 328, 1073
- The odds ratio - Altman DG & Bland JM BMJ (2000) 320, 1468

Sensitivity and Specificity

http://courses.ncssm.edu/math/Stat_Inst/Notes.htm

No author. North Carolina School of Science and Mathematics Statistics Leadership Institute Notes, July 1999. 3 pages.

Epidemiology in Practice: Case-Control Studies http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1706071/pdf/jceh_11_28_057.pdf Lewallen S & Courtright P (1998). Community Eye Health, 11, 57-58.

Standardization: a classic epidemiological method for the comparison of rates

<http://www.bvsde.paho.org/bvsea/fulltext/comparison.pdf>

[No authors listed], Epidemiological Bulletin (2002) 23: 9-12.

Age standardisation - an indigenous standard? <http://www.ete-online.com/content/pdf/1742-7622-4-3.pdf>

Robson B, Purdie G, Cram F & Simmonds S (2007) Emerging Themes in Epidemiology 4, 3,11 pages.

Why do we continue to use standardised mortality ratios for small area comparisons?

<http://jpubhealth.oxfordjournals.org/content/23/1/40.full.pdf>

Julious SA, Nicholl J & George S (2001) Journal of Public Health Medicine 23, 40-46.

ONLINE COURSES ON CENSORED DATA: SURVIVAL AND RELIABILITY

Applied Survival Analysis <http://www.amstat.org/chapters/northeasternillinois/pastevents/summer05.htm> http://www.amstat.org/chapters/northeasternillinois/pastevents/presentations/summer05_ibrahim_J.pdf

“Topics to be reviewed include non-parametric estimation of survival functions and construction of life tables, comparison of survival distributions, modeling survival data using the Cox proportional hazards regression model, model selection for survival analysis and assessing the proportional hazards assumption.”

Ibrahim, JG (2005) The 21st Annual Summer Workshop of the Northeastern Illinois Chapter of the American Statistical Association, 491 pages.

Statistical Reasoning I <http://ocw.jhsph.edu/index.cfm/go/viewCourse/course/StatisticalReasoning1/coursePage/index/>

- Module 4: Introduction to survival analysis - covers Kaplan-Meier estimate of survival function and log-rank test

Statistical Reasoning I & II <http://ocw.jhsph.edu/index.cfm/go/viewCourse/course/StatisticalReasoning2/coursePage/index/>

- Module 4: Survival analysis - covers Cox proportional hazards regression

ONLINE ARTICLES ON CENSORED DATA: SURVIVAL AND RELIABILITY

Survival analysis: A primer

<http://www.stat.berkeley.edu/~census/survival.pdf>

Freedman, DA (2008) Survival analysis: A primer. PDF Preprint. Later published as Freedman, DA (2008) The American Statistician 62, 110-119.

Survival analysis part I - basic concepts and first analyses

<http://researchonline.lshtm.ac.uk/3629/1/89-6601118a.pdf>

Clark TG, Bradburn MJ, Love SB & Altman DG (2003) British Journal of Cancer 89, 232-238.

Survival analysis part II: multivariate data analysis - an introduction to concepts and methods

<http://researchonline.lshtm.ac.uk/3630/1/89-6601119a.pdf>

Bradburn MJ, Clark TG, Love SB & Altman DG (2003) British Journal of Cancer 89, 431-436.

Survival analysis Part III: multivariate data analysis - choosing a model and assessing its adequacy and fit

<http://researchonline.lshtm.ac.uk/3631/1/89-6601120a.pdf>

Bradburn MJ, Clark TG, Love SB & Altman DG (2003) British Journal of Cancer 89 (4). pp. 605-611.

Survival analysis part IV - further concepts and methods in survival analysis [http://](http://researchonline.lshtm.ac.uk/3629/1/89-6601117a.pdf)

researchonline.lshtm.ac.uk/3629/1/89-6601117a.pdf

Clark TG, Bradburn MJ, Love SB & Altman DG (2003) British Journal of Cancer 89,781-786.

ONLINE COURSES ON MULTIVARIATE METHODS

STAT 505 Applied Multivariate Statistical Analysis <https://onlinecourses.science.psu.edu/stat505/>

- Lesson 2: Measures of central tendency, dispersion and association. Covers covariance and correlation matrices.
- Lesson 3: Linear combinations of random variables. Covers interval estimate of multivariate means.
- Lesson 4: Multivariate normal distribution
- Lesson 5: Sample mean vector and sample correlation and related inference problems. Covers correlation matrices.
- Lesson 7: Principal components analysis
- Lesson 10: Discriminant analysis
- Lesson 11: Inferences regarding the multivariate population mean. Covers Hotelling T-square test.
- Lesson 12: Cluster analysis

Clustering and classification methods for biologists <http://www.alanfielding.co.uk/multivar/index.htm>

- Principal components analysis <http://www.alanfielding.co.uk/multivar/pca.htm> •
- Discriminant analysis <http://www.alanfielding.co.uk/multivar/da.htm>
- Hierarchical cluster analysis <http://www.alanfielding.co.uk/multivar/ca.htm>
- K-means clustering <http://www.alanfielding.co.uk/multivar/kmeans.htm>

ONLINE ARTICLES ON MULTIVARIATE METHODS

Chapter 9 Cluster Analysis - download sample pages 237-284

A Concise Guide to Market Research by Sarstedt, Marko & Mooi, Erik

<http://www.springer.com/gb/book/9783642539640>

Chapter 23 Cluster Analysis

Chapter 25 Discriminant Analysis

Business Research Methods and Statistics Using SPSS by Burns, Robert & Burns, Richard

<http://www.uk.sagepub.com/burns/chapters.htm>

Principal component analysis <http://pubs.rsc.org/en/content/articlepdf/2014/ay/c3ay41907j>

Analytic Methods, 2014, 6, 2812 by Bro, Rasmus and Smilde, Age K "Principal component analysis is one of the most important and powerful methods in chemometrics as well as in a wealth of other areas. This paper provides a description of how to understand, use, and interpret principal component analysis. The paper focuses on the use of principal component analysis in typical chemometric areas but the results are generally applicable."

Introduction to Principal Components Analysis

<http://pubs.rsc.org/en/content/articlepdf/2014/ay/c3ay41907j>

PM&R, Volume 6, 275-278 by Sainani, Kristin L. "In this article I first provide examples that illustrate the utility of PCA and then delve into greater detail about how it works."