

Matilda Statistics Resource Portal

Matthew Sunderland & Tim Slade

2025-01-21

Contents

1	Welcome to the MSRP	5
1.1	Usage	6
2	Reproducible and replicable data analysis	7
2.1	Journal articles and textbooks	7
2.2	Websites	9
2.3	Online courses/workshops	9
3	Data visualisation	11
3.1	Journal articles and textbooks	11
3.2	Websites	12
3.3	Online courses and videos	14
4	Statistical Computing	17
4.1	SPSS	17
4.2	Stata	19
4.3	SAS	20
4.4	R	21
4.5	Mplus	23
4.6	Comparing programs	24
5	MLM and Latent growth models	27
5.1	Journal Articles and Textbooks	27
5.2	Websites	30

5.3	Online courses and videos	31
5.4	Example syntax	31
6	Causal Mediation Analysis	33
6.1	Journal Articles and Textbooks	33
6.2	Websites	35
6.3	Online courses and videos	35
6.4	Example syntax	36
7	Meta-analysis	37
7.1	Journal Articles and Textbooks	37
7.2	Online courses and videos	39
7.3	Software	39
8	Missing data analysis	41
8.1	Journal Articles and Textbooks	41
8.2	Websites	42
8.3	Online courses and videos	43
8.4	Software	44
9	Logistic regression	45
9.1	Journal Articles and Textbooks	45

Chapter 1

Welcome to the MSRP

A one-stop statistical shop for all your research needs



This book was written to provide easy access to the series of Matilda Centre Statistical resources. The chapters were assembled and annotated on a variety of topics tailored to the needs of researchers at the Matilda Centre for Research in Mental Health and Substance Use.

This book is a living and evolving document with regular additions made to each chapter as well as additional chapters based on further need. We encourage you to share this document widely and feel free to provide suggestions for additional resources, new chapters, or edits by contacting either Matthew Sunderland or Tim Slade.

1.1 Usage

You will find each of the individual topics in the left hand column. Within each chapter you will find several categories of resources including but not limited to textbooks, journal articles, example code, YouTube lectures and videos, and other useful websites or blogs.

This document was created using bookdown and the source code can be found on Github via clicking the “View book source” link at the bottom of the left hand column.

Chapter 2

Reproducible and replicable data analysis

This chapter provides resources to improve the reproducibility and replicability of your data analysis. In brief, these concepts represent the ability for anyone to take your analysis and be able to replicate and reproduce all your results. These concepts have always been important but have seen increasing attention given the replicability crisis in social psychology and science more broadly. Moreover, the open science movement and the ability to use online portals such as the OSF and GitHub have made these principles easier to implement. This list also contains useful guides on how to improve record keeping, organisation, coding, cleaning, and manipulating data while at the same time producing dynamic reports that track everything you do (and all the changes) during your data analysis. This list does rely heavily on using R software, but the same principles can be applied to other statistical software packages including Stata and SPSS. I would encourage more people to make the switch to R programming (it's not as hard as you might think). The main message is to use syntax/coding and annotate as much as possible! No more point-and-click or having to re-do tables in excel!!

2.1 Journal articles and textbooks

DeBruine, L., Barr, D. (2021). Data skills for reproducible research. Self-published. <https://psyteachr.github.io/reprores-v2/index.html#how-to-learn-data-skills>

- *An excellent online textbook on how to improve your data skills with a focus on tidy data, reproducible workflows, dynamic reports, and some*

more advanced topics. I recommend going over chapter 2 and chapter 6. The great thing about this book is that it has several accompanying YouTube videos outlining the concepts!

Epskamp, S. (2019). Reproducibility and Replicability in a fast-paced methodological world, *Advances in Methods and Practices in Psychological Science*, 2(2), 145-155.

- *Excellent overview of why we need reproducibility and replicability in data analysis. Provides hints and tips on how to improve reproducibility of results and selecting software to assist.*

Baumer, B., Centinkaya-Rundel, M, Bray, A., et al. (2014). R Markdown: Integrating a reproducible analysis tool into introductory statistics, *Technology Innovations in Statistics Education*, 8(1). <https://escholarship.org/uc/item/90b2f5xh>

- *A textbook covering the basics of R markdown and how it can be applied to generate reproducible statistics.*

De Jonge, E., van der Loo, M. (2013). An introduction to data cleaning with R. Statistics Netherlands, The Hague, NL. https://cran.r-project.org/doc/contrib/de_Jonge+van_der_Loo-Introduction_to_data_cleaning_with_R.pdf

- *A discussion paper outlining the basics of data cleaning in R. This paper is a little old now and might not have the most up to date methods (i.e. tidyverse).*

Buttrey, S., Whitaker, L. R. (2018). A data scientist's guide to acquiring, cleaning, and managing data in R. Wiley & Sons: Hoboken, NJ. (See PDF for copy).

- *A great textbook outlining the goals of data cleaning, reproducibility, basics in R, more advanced topics in R, and practical examples. I recommend going over Chapter 7 for clear examples.*

Collier, J. (2009). Using SPSS syntax a beginner's guide. SAGE: Los Angeles, CA. https://sydney.primo.exlibrisgroup.com/permalink/61USYD_INST/1c0ug48/alma991007515259705106

- *A textbook for those who want to move from point-and-click to programming using syntax in SPSS... but really if you are going to the effort to learn syntax in SPSS then you are probably better off starting fresh in R.*

2.2 Websites

<https://osf.io/>

- *For those unfamiliar with the Open Science Framework then please check out the above website. It provides a portal for store all your code, data, reports, and preprints of your work. All completely open and accessible. Please check the ethical requirements of data sharing before you share any data.*

<https://docs.github.com/en> and <https://docs.github.com/en/get-started/quickstart/hello-world>

- *This is an overview of GitHub. Not many people know that the Matilda Centre has a GitHub page (<https://github.com/Matilda-Centre>)! We encourage people to use GitHub to share and collaborate on analysis code. It is a great way to manage record keeping and stay on track with version control. If you want to join the Matilda Centre GitHub page then create your own profile and send me (Matt) an email with your username.*

<https://www.povertyactionlab.org/resource/data-cleaning-and-management>

- *Provides some useful examples and techniques for cleaning and managing data as well as ensuring proper version control and organisation.*

<https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf>

- *An R Markdown cheat sheet*

2.3 Online courses/workshops

<https://reproducible-analysis-workshop.readthedocs.io/en/latest/6.RMarkdown-knitr.html>

- *An online tutorial for using R to generate reproducible scientific analysis with a specific focus on R Markdown to generate dynamic reports.*

<https://bookdown.org/aschmi11/RESMHandbook/data-preparation-and-cleaning-in-r.html>

- *An online course focusing on basics in R, with good examples of using the Tidyverse to clean and manipulate data.*

<https://libguides.library.kent.edu/SPSS/home>

- *An online tutorial for working with data using SPSS. Focus on section 2 and section 3 for some nice examples of cleaning and manipulating data. It specifically draws links between the point-and-click methods of analysing data and the equivalent syntax to ease the transition.*

<https://www.youtube.com/watch?v=Cn-72tbRNFc>

- *Online lecture on designing a reproducible workflow with R and Github.*

Chapter 3

Data visualisation

This chapter provide resources for those who want to produce fancy figures and tables when communicating their results. Data visualisation is an important aspect of statistics and requires careful thought about how to appropriately and effectively present your results in a visually pleasing manner.

3.1 Journal articles and textbooks

Wilkinson, L. (2005). The grammar of graphics (2nd Edition). Springer.

- *Textbook that outlines foundation of graphing and is utilised in the R ggplot2 package. Available online: https://sydney.primo.exlibrisgroup.com/permalink/61USYD_INST/12rahnq/alma991017882639705106*

Wickham, H. (2016). Ggplot2: elegant graphics for data analysis (2nd Edition). Springer.

- *Guide and manual for the ggplot2 package for R. Very useful graphing package. Available online: https://sydney.primo.exlibrisgroup.com/permalink/61USYD_INST/1c0ug48/alma991014142669705106*

Winston, C. (2018). R graphics cookbook: practical recipes for visualising data (2nd Edition). O'Reilly Media.

- *Provide useful examples on how to visualise data and produce useful graphs. Primarily uses the ggplot2 package. Available online: https://sydney.primo.exlibrisgroup.com/permalink/61USYD_INST/2rsddf/cdi_askewsholts_vlebooks_9781491978573*

Healy, K. (2019). Data visualization: A practical introduction. Princeton University Press.

- *Introductory textbook on graphing and data viz. Focuses on ggplot2 and the tidyverse.*

Healy, K., & Moody, J. (2014). Data visualisation in sociology. *Annu. Rev. Sociol.* 40: 150-128.

- *Article outlining data viz specific to sociology. Good place to see how to work towards a consistent standard for graphing sociological insights.*

Ola, O., & Sedig, K. (2016). Beyond simple charts: design of visualisations for big health data. *Online J Public Health Inform* 8(3): e195.

- *Provides a framework for designing graphs associated with “big” data.*

Slaughter, S. J., & Delwiche, L. D. (2010). Using PROC SGPLOT for quick high-quality graphs. *SAS Global Forum*.

- *SUGI paper on graphing with SAS... if that’s your thing. <https://support.sas.com/resources/papers/proceedings10/154-2010.pdf>*

Mitchell, M. N. (2012). A visual guide to stata graphics. Stata Press.

- *Introductory textbook for graphing using STATA. Also has accompanying website <https://www.stata.com/support/faqs/graphics/gph/stata-graphs/>*

3.2 Websites

<https://www.jscarlton.net/post/2015-10-24visualizinglogistic/>

- *Demonstrates one way to present odds ratios and CIs graphically.*

<https://www.r-graph-gallery.com/>

- *Provides lots of inspiration and example code for graphing in R.*

<https://www.tableau.com/>

- *Online platform for data graphing and visualization.*

<https://dreamrs.github.io/esquisse/index.html>

- *Package that allows you to interactively design graphs quickly in R.*

<https://www.data-to-viz.com/>

- *Awesome webpage that provides a nice decision tree around what graph to use for the type of data you have.*

<https://seaborn.pydata.org/>

- *This is a large data visualisation library for Python.*

<https://rstudio.github.io/leaflet/> or <https://leafletjs.com/>

- *Open-source JavaScript library for interactive maps.*

<https://bookdown.org/ndphillips/YaRrr/pirateplot.html>

- *Example of a pirate plot that provides more usable information on a single graph than the traditional box plot.*

<https://stats.idre.ucla.edu/spss/seminars/spssgraphics/graphics-in-spss/>

- *Graphing in SPSS... if that's your thing.*

<https://stats.idre.ucla.edu/sas/modules/graphing-data-in-sas/>

- *More info on graphing in SAS.*

<https://stats.idre.ucla.edu/stata/modules/graph8/intro/introduction-to-graphs-in-stata/>

- *Graphing in Stata... if that's your thing.*

<https://data.library.virginia.edu/getting-started-with-shiny/>

- *Website that provides an introduction to Shiny (a package that lets you create interactive web apps to visualise data).*

<https://shiny.rstudio.com/images/shiny-cheatsheet.pdf>

- *Shiny cheat sheet*

<https://www.rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf>

- *Data viz cheat sheet in ggplot2*

<https://plotly.com/>

- *Similar to shiny and lets you create web apps to vizualise data. Can use Python, R codes.*

<https://plotly.com/graphing-libraries/>

- *Extensive graph library in Python, R, and Javascript. Provides lots of inspiration.*

[https://www.gapminder.org/tools/#\\$chart-type=bubbles](https://www.gapminder.org/tools/#$chart-type=bubbles)

- *Fun online tool to explore how powerful data viz can be.*

<https://developers.google.com/chart>

- *Google charts. This is more for developing interactive charts for web-sites/mobile devices.*

3.3 Online courses and videos

<https://www.andrew.cmu.edu/user/achoulde/94842/lectures/lecture13/lecture13-94842.html>

- *Lecture material to introduce interactive graphics and shiny.*

<https://www.youtube.com/watch?v=49fADBfcDD4>

- *Live tutorial on intro to data viz with R and ggplot2*

<https://www.youtube.com/watch?v=h29g21z0a68>

- *Another lecture/workshop on ggplot2. There are two parts to this workshop.*

<https://www.youtube.com/watch?v=zzXCkYR84M0>

- *Intermediate lecture on ggplot2 offered by UQ.*

<https://www.youtube.com/watch?v=TPMlZxRRaBQ>

- *Tableau crash course.*

<https://www.youtube.com/watch?v=Gyrfsrd4zK0>

- *How to create a web application in Shiny.*

<https://online-learning.harvard.edu/course/data-science-visualization?delta=2>

- *Online data viz course offered by Harvard via EdX.*

Chapter 4

Statistical Computing

In this chapter we cover some of the more commonly used statistical programs. This chapter provides introductory materials when starting to use a particular software package. In terms of getting access to these programs, you should be able to download SPSS and SAS from the University of Sydney IT services. Stata requires an individual licence, or it can be used for free remotely via the Usyd Citrix Workspace App. R and Rstudio (highly recommended) can be downloaded for free along with the many R packages (<https://www.r-project.org/> <https://rstudio.com/>). Mplus requires a license and we currently have a couple of copies of Mplus installed on the Matilda Stats computers (i.e., in Matt's office and on the Statistics's laptop).

4.1 SPSS

<https://www.spss-tutorials.com/>

- *Website that provides online tutorials on a range of analyses in SPSS as well as the basic overview of using the program.*

<https://libguides.library.kent.edu/SPSS/home>

- *Online tutorials on SPSS offered by Kent State University. Covers getting started, working with data, exploring, and analysing data.*

<https://lo.unisa.edu.au/mod/book/view.php?id=646443&chapterid=106603>

- *Online tutorial offered by University of SA on SPSS. Quite good for getting started with Syntax.*

<https://stats.idre.ucla.edu/spss/modules/>

- *Excellent website from UCLA on the basics of SPSS, including fundamentals, reading raw data, and data management.*

https://milton-the-cat.rocks/dsus_oditi.html

- *An awesome series of videos on the use of SPSS by Andy Field taken from his textbook “Discovering Statistics Using IBM SPSS Statistics”. I also would advise getting a copy of the textbook. <https://amzn.asia/d/0CC4lWt>. In fact Andy Field’s main website is amazing and would suggest everyone check it out <https://www.discoveringstatistics.com/>*

<https://cce.sydney.edu.au/course/SPSS>

- *Formal SPSS beginner course offered online via the Usyd Centre for Continuing education. Cost involved.*

https://www.youtube.com/playlist?list=PL7Opqhbc9Wr0GG4IZeLVO8px_e1tYYP6g

- *A YouTube series on SPSS hacks.*

<https://www.dummies.com/education/math/statistics/spss-for-dummies-cheat-sheet/>

- *SPSS syntax cheat sheet from the SPSS for dummies book.*

Pallant, J., (2020). SPSS survival manual: A step by step guide to data analysis using IBM SPSS (7th Edition). Routledge:

- *A very handy textbook if you can get a copy. Companion website here <https://routledgetextbooks.com/textbooks/9781760875534/>*

Levesque, R., (2005). SPSS programming and data management, 2nd Edition. SPSS Inc: Chicago, IL.

- *Textbook available here https://spsstools.net/static/spss-programming-book/SPSS_Programming_and_Data_Management_2nd_edition.pdf. Offers a large amount of detail in using SPSS from basics to advanced programming and macros. It is also quite helpful if you want to migrate to SAS from SPSS... but can be a little dry.*

4.2 Stata

<https://www.stata.com/bookstore/>

- *This link allows you to browse all the available textbooks on Stata offered via the official website.*

<https://www.stata.com/bookstore/stata-journal/>

- *Link to the official Stata Journal that provides papers on a range of statistical topics using Stata.*

<https://stats.idre.ucla.edu/stata/>

- *UCLA website on the use of Stata some excellent resources including data examples, annotated output, textbook examples, and learning modules.*

<https://data.princeton.edu/stata/>

- *Stata tutorial available from Princeton University. Includes an introduction to Stata, data management, graphics, and programming.*

<https://www.stata.com/support/faqs/>

- *Quite a useful FAQ website offered by Stata that provides answers to questions on statistical analysis, programming, graphics, data management, etc.*

<http://www.stat.columbia.edu/~tzheng/teaching/w1111/startingstata.htm>

- *Very brief overview on getting started with Stata offered by Columbia University.*

<https://www.stata.com/bookstore/handbook-statistical-analyses-stata/>

- *This book is a little old but it provides some good examples of how to carry out different data analysis procedures (e.g. logistic regression, survival analysis, GEE) in Stata.*

4.3 SAS

<https://online.stat.psu.edu/stat480/lesson/1>

- *Penn State University online lesson on Getting started in SAS.*

<https://www.youtube.com/channel/UCWOfmTlbeesYiDJNflqsWQA>

- *This is the SAS Users YouTube website. Official SAS trainers provide a range of video tutorials on the basics of programming and analysis using SAS. Also covers additional topics like SAS Enterprise Guide and other add-ins.*

<https://stats.idre.ucla.edu/sas/>

- *Again, excellent website from UCLA on basics of using SAS as well as an extensive list of data analysis examples, annotated outcome, textbook examples, and other useful resources. I would recommend look at the SAS Library link <https://stats.idre.ucla.edu/sas/library/> as this provides an extensive range of resources.*

https://www.sas.com/store/books/categories/getting-started/cBooks-cbooks_categories-cbooks_categories_3-p1.html

- *Official list of beginner textbooks from SAS. I would recommend “Fundamentals of Programming in SAS” and “The Little SAS Book”. You should be able to find these textbooks in the Usyd library if you search.*

https://www.sas.com/store/books/categories/analytics-statistics/cBooks-cbooks_categories-cbooks_categories_18-p1.html

- *More textbooks officially endorsed by SAS but this time focused on analysis and various types of statistical procedures.*

https://www.lexjansen.com/cgi-bin/xsl_transform.php?x=sbt&c=sugi

- *This website provides Beginning tutorials from the SAS Global Forum (previously called SUGI). Attendees of the forum must provide papers if they are presenting and many papers demonstrate how to conduct specific analyses or tips on how to improve your SAS skills.*

<https://www.lesahoffman.com/PSYC930SAS/index.html>

- *Some lecture materials on “Making friends with SAS” offered by Lesa Hoffman.*

<https://support.sas.com/edu/courses.html?ctry=AU>

- *This is a link to all courses offered by SAS, some of free but the vast majority require payment and can be quite expensive but very good if you can a chance to attend one.*

DiMaggio, C., (2013). SAS for Epidemiologists: Applications and Methods. New York, NY: Springer.

- *A comprehensive textbook mostly designed for students studying public health and epidemiology. Covers topics such as data cleaning, programming, categorical data analysis, ANOVA, linear regression. The textbook is available online via Usyd library https://sydney.primo.exlibrisgroup.com/permalink/61USYD_INST/1c0ug48/alma991019154709705106*

4.4 R

Important note: R is a large and flexible program with a tonne of free resources available. This is just the tip of the iceberg and if you are wanting to use R then Google becomes your best friend. It is easy to get overwhelmed, so I suggest you start working on smaller projects and tackle tasks and topics that are suitable for that project one at a time.

<https://www.r-bloggers.com/2017/03/the-5-most-effective-ways-to-learn-r/>

- *A blog post on the most effective ways to learn R*

<https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf>

-An introductory manual for R offered by the R core team.

<http://www.r-tutor.com/r-introduction>

- *Website that provides a series of introductory tutorials on R including basic data types, vectors, matrices, lists, data frames, and some beginning statistics.*

<https://rstudio.com/resources/webinars/a-gentle-introduction-to-tidy-statistics-in-r/>

- *A gentle introduction to Tidy Statistics.*

<https://rstudio.com/collections/rstudio-essentials/>

- *This provides several webinars on the use of RStudio. It is highly recommended that you use R via the Rstudio software. This is a more useable interface than the base R software.*

<https://rstudio.com/resources/webinars/getting-started-with-r-markdown/>

- *This video looks at getting started with RMarkdown. This allows you to generate reproducible reports using R code.*

<https://education.rstudio.com/learn/beginner/>

- *Beginner resources that are offered by RStudio and include getting started guides, cheat sheets, textbooks, tutorials, etc. Some recommended textbooks include: R for Data science, R Cookbook, and the R Graphics Cookbook. I also highly recommend looking through the cheat sheets and picking out ones that you will most likely refer to often.*

<https://www.statmethods.net/index.html>

- *In depth website that provides tutorials for beginners in using R, including statistics, advanced statistics, graphs, and advanced graphs.*

<https://www.youtube.com/user/marinstatlectures>

- *A very informative YouTube channel dedicated to statistics predominately in R*

<https://www.r-graph-gallery.com/>

- *A useful website dedicated to graphing in R. It provides lots of examples for many different chart types, including graphing distributions, correlation, ranking, evolution, mapping, flow, and animated graphs.*

<https://www.r-bloggers.com/>

- *A link to many different blogs on R including new books, how to perform various statistical procedures in R, and how to improve your programming.*

<https://journal.r-project.org/>

- *Link to the R journal that includes papers in various R packages.*

<https://intersect.org.au/training/courses/>

- *Online Zoom lessons in R, Python and other software offered free to USyd staff. Ranges from novice to advanced topics like machine learning. I have done a course here and found it really helpful. It is great to be able to ask little questions that arise in real-time.*

<https://www.sydney.edu.au/research/facilities/sydney-informatics-hub/workshops-and-training/training-calendar.html>

- *This is a calendar for the Sydney informatics hub workshops. Most are free, but some have a fee (~\$500) Note some of these link to courses run by intersect.*

<https://r-graphics.org/>

- *A link to the online version of the book “R Graphics Cookbook” which provides practical “recipes” for visualising data*

4.5 Mplus

<https://stats.idre.ucla.edu/mplus/>

- *Once more, the UCLA website this time focusing on Mplus. Provides seminars, common questions, analysis examples, textbook examples, annotated outcome.*

<http://www.statmodel.com/ug excerpts.shtml>

- *Link to the official Mplus user’s guide and language addendums. This site also contains all the example input programs for the examples offered in each chapter of the user’s guide.*

<http://www.statmodel.com/cgi-bin/discus/discus.cgi>

- *Official Mplus discussion forum that can sometimes be useful if you want to look for something specific. You can only post of the forum if you have an active Mplus license.*

<https://www.youtube.com/c/MplusVideos>

- *The official Mplus YouTube website that contains all the lectures and short courses on using Mplus for various statistical procedures.*

<https://www.youtube.com/playlist?list=PL8aoPJZQes3qZbWeho7WFNjgpRtLU3KHs>

- *YouTube playlist of Michael Zyphur's Mplus workshop conducted over 5 days at the University of Melbourne. Covers a range of different topics... mostly more advanced.*

<https://www.youtube.com/user/ltolandky/videos>

- *YouTube channel by Michael Toland. Provides several videos on introduction to Mplus, as well as specific topics on path analysis, confirmatory factor analysis, bifactor analysis, etc.*

<https://global.oup.com/us/companion.websites/9780195367621/pdf/MplusQuickGuide2015.pdf>

- *Introductory guide in using Mplus. Provides an overview of the program, the language, some various models, and using the Mplus language generator.*

<https://cran.r-project.org/web/packages/MplusAutomation/vignettes/Vignette.pdf>

- *Highlights a very useful package in R that works with Mplus called MplusAutomation. This PDF provides some basic examples on how to start using Mplus with R. It takes advantage of the data manipulation and graphing power of R but the statistical power of the Mplus engine.*

4.6 Comparing programs

Important note: there is no right or wrong package to use if it can do the task that you require. Our advice would be to learn something that you feel most comfortable with using. Many people already have experience with SPSS and you can continue to use it, other people want to start fresh and might begin learning R or Mplus. The following links might help guide you in terms of which software package to start using.

<https://guides.nyu.edu/quant/statsoft>

- *NYU comparing of different programs providing basic pros and cons of each. Covers SPSS, Stata, SAS, R and a couple of additional programs not mentioned here.*

<https://www.r-bloggers.com/2019/07/whats-the-best-statistical-software-a-comparison-of-r-python-sas-spss-and-stata/>

- *Another blog post on the different merits of each statistical program. Compares R, python, SAS, SPSS, and Stata*

<https://lo.unisa.edu.au/mod/book/view.php?id=631718>

- *Comparison of programs by University of SA.*

Chapter 5

MLM and Latent growth models

This chapter provides resources on multilevel models (MLM) and latent growth models for the analysis of longitudinal data. Both MLM and latent growth models are similar but different frameworks to analyse change over time. Both have strengths and weaknesses and both frameworks might be more or less suitable depending on the specific research question and the type of data that you want to analyse. Whilst MLM can be used for other types of clustered or multi-level structured data, we focus here on the use of MLM for repeated measures or longitudinal analysis.

5.1 Journal Articles and Textbooks

Hoffman, L. (2015). Longitudinal analysis: modelling within-person fluctuation and change. Routledge.

- *An excellent textbook that provides easy to understand text and examples on the analysis of longitudinal data. Written primarily for non-mathematical readers. It focuses on multilevel models for longitudinal data, looking at unconditional and conditional models, time-invariant and time-varying predictors, and more advanced applications. I have got a copy of the book if anyone wants to borrow it or it can be found online via the Usyd library website. https://sydney.primo.exlibrisgroup.com/permalink/61USYD_INST/2rsddf/cdi_askewsholts_vlebooks_9781317591092*

Duncan, T. E., Duncan, S.C., and Strycker, L. A. (2006) An introduction to latent variable growth curve modelling: concepts, issues, and application (2nd ed.) Routledge.

- *Another great textbook but focused more on using latent growth models for longitudinal data (also sometimes referred to as a structural equation modelling or multivariate approach to longitudinal data). Contains a range of topics including growth models for non-continuous data, growth mixture models, piecewise models, how to handle missing data, etc. This book is also available online via Usyd website. https://sydney.primo.exlibrisgroup.com/permalink/61USYD_INST/2rsddf/cdi_askewsholts_vlebooks_9780203879962*

Twisk, J. W. R. (2013). *Applied longitudinal data analysis for epidemiology (a practical guide)* (2nd ed.) Cambridge University Press.

- *A very useful textbook that provides a general overview of some of the most important techniques for longitudinal data analysis, particularly from an epidemiological data perspective. It covers multiple types of data (continuous, binary, ordinal, count), the analysis of experimental studies, and how to handle missing data. The book is available online via Usyd website. https://sydney.primo.exlibrisgroup.com/permalink/61USYD_INST/2rsddf/cdi_askewsholts_vlebooks_9781107065352*

Twisk, J. W. R. (2021) *Analysis of data from randomised controlled trials: A practical guide*. Springer.

- *Another useful textbook by Twisk that provides a general overview of analysis of RCTs. There are chapters that focus on using multilevel models for the analysis of RCTs with more than one follow-up measurement. The book also provides an overview of cluster RCTs and dichotomous outcomes that are commonly used in RCTs at Matilda. <https://link.springer.com/book/10.1007/978-3-030-81865-4>*

Singer, J. D., & Willett, J. B. (2003) *Applied longitudinal data analysis: modelling change and event occurrence*. Oxford University Press.

- *Another textbook focusing on longitudinal data analysis. This book contains material on multilevel models but also on survival models. Lots of examples are provided. The book is available online via the Usyd library website. https://sydney.primo.exlibrisgroup.com/permalink/61USYD_INST/2rsddf/cdi_proquest_ebookcentral_EBC3054153*

Grimm, K. J., Ram, N., & Estabrook, R. (2016). *Growth modelling: structural equation and multilevel modelling approaches*. Guilford Press.

- *Just in case you needed another textbook on growth models. This provides many real-world examples using longitudinal data. Also available online via Usyd library website https://sydney.primo.exlibrisgroup.com/permalink/61USYD_INST/1c0ug48/alma991013485709705106*

Chih-Ping Chou, Peter M. Bentler & Mary Ann Pentz (1998) Comparisons of two statistical approaches to study growth curves: The multilevel model and the latent curve analysis, *Structural Equation Modeling: A Multidisciplinary Journal*, 5:3, 247-266, DOI: <https://10.1080/10705519809540104>

- *A useful paper that describes differences between multilevel modelling framework and the latent growth curve framework (sometimes called structural equation modelling or multivariate approach). Highlights the pros and cons of both frameworks depending on the needs of the analysis and the data you have.*

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve Frequently Asked Questions About Growth Curve Modeling. *Journal of cognition and development : official journal of the Cognitive Development Society*, 11(2), 121–136. <https://doi.org/10.1080/15248371003699969>

- *Nice overview article of latent growth curve models. Provides an introduction, definitions, data requirements, missing data, different growth shapes, model fit, etc.*

Hesser, H. (2015). Modeling individual differences in randomized experiments using growth models: Recommendations for design, statistical analysis and reporting of results of internet interventions. *Internet Interventions*, 2(2), 110–120. <https://doi.org/10.1016/J.INVENT.2015.02.003>

- *Another introduction but specifically focused on RCTs and internet interventions. Particularly relevant to the work done at Matilda.*

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology*, 48(1), 85–112. <https://doi.org/10.1016/j.jsp.2009.09.002>

- *A paper is that more generally focused on multilevel model framework. Can be applied to nested cross-sectional data or longitudinal data. Provides examples of both.*

Feingold, A. (2013). A regression framework for effect size assessments in longitudinal modelling of group differences. *Review of General Psychology*, 17(1), 111-121. DOI: <https://doi.org/10.1037/a0030048>

- *A paper focused on the use of estimating effect sizes associated with longitudinal growth models. Particularly relevant for identifying the magnitude of group differences (e.g., intervention data).*

Ram, N. & Grimm, K. (2007). Using simple and complex growth models to articulate developmental change: matching theory to method. *International Journal of Behavioural Development*, 31(4), 303-316. <https://doi.org/10.1177/0165025407077751>

- *Another more focused paper, this time looking at non-linear models of change, e.g. quadratic, latent bias, exponential, multiphasic.*

McNeish, D., Matta, T. (2018). Differentiating between mixed-effects and latent-curve approaches to growth modeling. *Behav Res* 50, 1398–1414. <https://doi.org/10.3758/s13428-017-0976-5>

- *An excellent article outlining the similarities and differences between MLM and latent growth models as well as discussion of the strengths and weaknesses and how to select the best framework for you research question/data.*

5.2 Websites

<https://www.pilesofvariance.com/>

- *This website accompanies the Hoffman textbook. It provides several examples in different statistical programs (SAS, SPSS, Stata, MPlus) linked to examples provided in the textbook.*

<https://hedeker.people.uic.edu/ml.html>

- *Don Hedeker's website that provides a range of examples in SPSS and SAS as well as different types of data (continuous, dichotomous, ordinal, nominal). They also provide a guide for estimating sample sizes for longitudinal designs.*

<https://sites.google.com/site/longitudinalmethods/home?authuser=0>

- *Kevin Grimm's lab website. Contains a list of publications and downloads for example syntax and scripts.*

<http://www.bristol.ac.uk/cmm/>

- *The centre for multilevel modelling at the University of Bristol provides a range of resources on multilevel modelling in general (not just applied to longitudinal data).*

5.3 Online courses and videos

<https://youtube.com/playlist?list=PLQGe6zcSJT0VxMZUN6DBuhIoCRZNoA2Vz>

- *A nine-part YouTube tutorial on growth curve modelling by Curran-Bauer Analytics. Covers the basics and the differences between longitudinal analysis using multilevel modelling vs structural equation modelling.*

https://www.lesahoffman.com/PSQF7375_Longitudinal_Spring2019/index.html

- *Video lecture series conducted by Lesa Hoffman as part of her course at University of Iowa.*

<https://youtu.be/2iQJCLiyl8>

- *Introductory and intermediate growth models using Mplus. This lecture is part of the Mplus short course conducted at Johns Hopkins University. Lecture notes available here <http://www.statmodel.com/download/Topic%203.pdf>*

<https://youtu.be/cr9RpSgRYVw>

- *A three-part series explaining how to analyse longitudinal data via multi-level models in R and RStudio.*

<https://stats.idre.ucla.edu/other/mult-pkg/seminars/mlm-longitudinal/>

- *Online notes for a seminar based on Singer and Willett textbook. Focuses on comparing different analytic strategies, multilevel level models, and treatment effects over time.*

5.4 Example syntax

https://stats.idre.ucla.edu/mplus/output/lgcm_mlm/

- *Annotated Mplus programs on latent growth and multilevel models. Provides four examples: two growth and two multilevel models. It also draws comparisons to other programs like HLM and Stata.*

<https://hedeker.people.uic.edu/ml.html>

- *Contains multiple example syntax and programs for SAS and SPSS.*

<https://stats.idre.ucla.edu/other/examples/alda/>

- *Example syntax from Singer and Willett textbook. Sample data is available to download along with programs in Mplus, SAS, Stata, R, and SPSS.*

<https://sites.google.com/site/longitudinalmethods/downloads?authuser=0>

- *Contains many example files and scripts from Mplus, SAS, OpenMX, etc. The examples come from papers published by Kevin Grimm.*

Chapter 6

Causal Mediation Analysis

This chapter provides resources that outline causal mediation analysis and the potential outcomes framework. Specifically, you will find resources defining and describing causal mediation analysis, how it differs from traditional mediation approach, and practical guidelines for conducting an analysis and how best to interpret the results.

6.1 Journal Articles and Textbooks

VanderWeele, T. (2015). *Explanation in causal inference: methods for mediation and interaction*. Oxford University Press.

- *Pretty much “the textbook” on causal mediation analysis written by one of the leading experts in the field. Covers many topics including sensitivity analysis, multiple mediators, time varying exposures, etc. The book is available to view online via the Usyd library.*

Muthén, B. O., Muthén, L. K., & Asparouhov, T. (2017). *Regression and mediation analysis using Mplus*. Los Angeles, CA: Muthén & Muthén.

- *An excellent textbook providing an introduction and overview of causal mediation analysis and real-data examples using Mplus software. See Chapter 4 for introduction to causal mediation analysis and Chapter 8 for causal mediation specifically for binary outcomes and continuous mediators (what we often are faced with in terms of data at Matilda). I have a copy of the book if anyone wants to borrow it.*

Muthén, B. & Asparouhov T. (2015). Causal effects in mediation modeling: An introduction with applications to latent variables. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(1), 12-23. DOI:10.1080/10705511.2014.935843

- *Introduces causal effect models and then offers applications with latent variables as the mediators or outcomes. Provides example Mplus scripts for mediated moderation analysis and binary outcomes. Alternative paper on similar topic but more detail found here <http://www.statmodel.com/download/causalmediation.pdf>*

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological methods*, 15(4), 309.

- *Article that provides and overview of causal mediation analysis and a general framework for analysis rather than use of linear structural equation models. The article is a little dense but provides some useful examples and software options.*

Rijnhart, J. M., Valente, M. J., MacKinnon, D. P., Twisk, J. W R., Heymans, M. W. (2020). The use of traditional and causal estimators for mediation models with a binary outcome and exposure-mediator interaction. *Structural Equation Modeling*, DOI: 10.1080/10705511.2020.1811709

- *Article that provides and overview of the distinction people traditional mediation and causal mediation and provides real-data examples, analytic comparisons, and simulation study to demonstrate differences and similarities between the two approaches. Concludes that causal mediation is the preferred method. Includes example codes for Stata.*

Feingold A, MacKinnon DP, Capaldi DM. Mediation analysis with binary outcomes: Direct and indirect effects of pro-alcohol influences on alcohol use disorders. *Addictive Behaviors*. 2019 Jul;94:26-35. DOI: 10.1016/j.addbeh.2018.12.018.

- *Articles that provides and overview of causal mediation specifically for binary outcomes and provides useful templates for presentation of results and interpretation of odds ratios.*

Valente, M. J., Rijnhart, J. J., Smyth, H. L., Muniz, F. B., & MacKinnon, D. P. (2020). Causal Mediation Programs in R, M plus, SAS, SPSS, and Stata. *Structural Equation Modeling*. DOI: 10.1080/10705511.2020.1777133

- *Article provides an overview of the various computer programs to conduct causal mediation analysis. The article draws comparisons between the programs and highlights strengths and weaknesses of each. Finds that many software packages are the same but Mplus and R are better when handling missing data. Mplus is preferred when any of the variables represent latent variables.*

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., Imai, K. (2014). Mediation: R package for causal mediation analysis. *Journal of Statistical Software*, 59(5).

- *Provides an overview and description of an R package to conduct causal mediation analysis using model-based and design-based approaches as described by Imai.*

6.2 Websites

<https://www.publichealth.columbia.edu/research/population-health-methods/causal-mediation>

- *Website from Columbia University Mailman School of Public health. Provides a brief overview and a list of additional readings and potential courses.*

<https://www.hsph.harvard.edu/tyler-vanderweele/tools-and-tutorials/>

- *Tyler VanderWeele website that provides additional tools and tutorials for conducting sensitivity analysis for unmeasured confounding (e-values), tools for power analysis, and links to online courses.*

<https://imai.fas.harvard.edu/projects/mechanisms.html>

- *Kosuke Imai's website providing list of publications and links to presentations and statistical software to conduct mediation analysis using Imai's approach in R.*

6.3 Online courses and videos

<https://www.youtube.com/watch?v=EI5y6pV87-Q> <https://www.youtube.com/watch?v=WyqdGxsnR5w> <https://www.youtube.com/watch?v=CNdan-6gNIY> <https://www.youtube.com/watch?v=WFMN1828648>

- *Parts 1-4 of an online lecture series presented by Tyler VanderWeele on causal mediation analysis. Provides an overview and covers topics such as binary outcomes, sensitivity analysis, and causal mediation in survival data.*

<https://www.youtube.com/watch?v=1EQa4z-509Y>

- *Lecture conducted by Bengt Muthen from Mplus on causal mediation analysis. Provides a very accessible overview of the method. https://www.youtube.com/playlist?list=PLasC_CrKi-FVLmDDnVk0FMgl1f04J73ck*
- *Full playlist from the Mplus short course on regression and mediation where the above lecture was taken from. Covers more topics such as linear regression, count variable modelling, traditional mediation analysis, sensitivity analysis, Bayesian analysis, and missing data analysis.*

https://www.youtube.com/watch?v=9j_HWkrSxzI

- *Brief overview and introduction on the potential outcomes framework and counterfactuals.*

<https://statisticalhorizons.com/seminars/public-seminars>

- *This organisation offers remote seminars on causal mediation analysis via zoom on a regular basis. It might be useful to check when the next seminar is on from time to time. There is a cost involved.*

6.4 Example syntax

<http://www.statmodel.com/examples/penn.shtml#extendSEM>

- *Example mplus scripts from the paper “Applications of causally defined direct and indirect effects in mediation analysis using SEM in Mplus”*

<http://www.statmodel.com/mplusbook/chapter8.shtml>

- *Example mplus scripts from the textbook Chapter 8 focusing on mediation of binary and count variables. Good examples include that for Table 8.9, Table 8.22, Table 8.2 and Table 8.5.*

<https://www.tandfonline.com/doi/full/10.1080/10705511.2020.1811709>

- *See supplementary material from Rijnhart et al., (2020) paper for example Stata codes for running causal mediation models with binary outcomes.*

Chapter 7

Meta-analysis

This chapter provides resources that outline basic meta-analysis to summary data obtained from systematic reviews as well as more complex meta-analytic techniques such as network meta-analysis and structural equation modeling meta-analysis. There are several useful textbooks all available online that comprehensively outline the steps required for various meta-analyses. The majority of resources utilise free software packages in R to conduct meta-analyses and present results, however there are several specific software programs for meta-analysis outlined in the software section.

7.1 Journal Articles and Textbooks

Hedges, L. V. & Olkin, I. (2014) Statistical Method for Meta-analysis. Elsevier Science & Technology.

- *This is an older textbook but still very helpful. It covers the main methodological approaches for meta-analysis and provides examples in statistical packages. This book is available to view online via the Usyd library https://sydney.primo.exlibrisgroup.com/permalink/61USYD_INST/1c0ug48/alma991006161679705106*

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009) Introduction to Meta-analysis. John Wiley & Sons, Ltd.

- *Another fantastic starting point when learning meta-analysis. It provides a clear introduction to meta-analysis and covers the major concepts as well as provides information about identifying and avoiding common mistakes. This book is available to view online via the Usyd library*

https://sydney.primo.exlibrisgroup.com/permalink/61USYD_INST/1c0ug48/alma991002771109705106

Harrer, M., Cuijpers, P., Furukawa, T.A., & Ebert, D.D. (2021). *Doing Meta-Analysis with R: A Hands-On Guide*. Boca Raton, FL and London: Chapman & Hall/CRC Press.

- *Another very helpful and user-friendly online textbook on conducting meta-analysis with R. They cover basic topics as well as some more advanced topics like structural equation modelling meta-analysis, network meta-analysis, and Bayesian meta-analysis. One of the cooler things they have is an effect size calculator and conversion in section 17. Highly recommended resource.* https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/

Field, A. P., & Gillett, R. (2010). How do to a meta-analysis. *British Journal of Mathematical and Statistical Psychology*, 63(3), 665-694.

- *A great introductory paper on how to do a meta-analysis and information on how to conduct them in SPSS and R.* <https://doi.org/10.1348/000711010X502733>

Paul, J., & Barari, M. (2022) Meta-analysis and traditional systematic literature reviews – what, why, when, where, and how? *Psychology and Marketing*, 39(6), 1099-1115.

- *Another introductory paper on meta-analysis and how they differ from traditional systematic reviews. There is some information more specific to business research that can be ignored.*

Rouse, B., Chaimani, A., & Li, T. (2017). Network meta-analysis: an introduction for clinicians. *Internal and Emergency Medicine*, 12(1), 103-111.

- *An introduction to a relatively new technique of meta-analysis that compares multiple treatments from RCTs via the use of network analysis. Network meta-analysis is probably a large enough topic on its own so this might form its own separate list if people are interested.*

Fernandez-Castilla, B., & Van den Nootgate, W. (2023). Network meta-analysis in psychology and educational sciences: A systematic review of their characteristics. *Behaviour Research Methods*, 55, 2093-2108

- *Another introductory paper on network meta-analysis but more tailored to psychological and educational research. Provides some useful warnings about what not to do when reporting a network meta-analysis.*

Deeks JJ, Higgins JPT, Altman DG (editors). Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated August 2023). Cochrane, 2023. www.training.cochrane.org/handbook

- *Online textbook from the Cochrane handbook of systematic reviews. Quite trustworthy and provides most reliable guidelines.*

Schwarzer, G., Carpenter, J. R., & Rucker, G. (2015). Meta-analysis with R. Springer Links.

- *A practical textbook outlining all the steps required to conduct a meta-analysis using the R package meta. <https://github.com/guido-s/meta>*

7.2 Online courses and videos

<https://www.youtube.com/watch?v=lO6s8dFBmPc>

- *A brief introduction to meta-analysis by Daniel Lakens.*

<https://www.youtube.com/watch?v=IH4VZMTEZSc>

- *A very accessible step-by-step guide to conducting a meta-analysis using R by Daniel Quintana. Be sure to check out some of his other YouTube videos on meta-analysis and reproducible research.*

<https://www.youtube.com/watch?v=Hiu3rF5ir1U>

- *1.5hr workshop on conducting open and reproducible meta-analyses.*

7.3 Software

Comprehensive meta analysis <https://meta-analysis.com/>

- *This is a software package specifically designed to meta-analysis and provides a very user friendly and systematic approach to meta-analysis. It has a wide range of support options and a long list of tutorials covering major types of meta-analysis. There is a cost to it but there are academic and student prices available. My own opinion is that if you are confident using R or another package then there is probably not much need for CMA.*

RevMan <https://training.cochrane.org/online-learning/core-software/revman>

- *This program is provided by Cochrane to assist with systematic reviews and can be used to conduct meta-analysis. It does require a subscription to use and might be useful if you want to run a systematic review and meta-analysis all within the same program.*

Metafor (R package) <https://www.metafor-project.org/doku.php/metafor>

- *One of the free and open sources R packages to conduct meta-analysis. It requires some knowledge of R but there are several examples and tutorials available to conduct the analysis using metafor.*

Chapter 8

Missing data analysis

Unless you manage to get perfect retention rates or are able to achieve 100% response rates on a survey, you will at some stage need to conduct a missing data analysis. These types of analysis can be quite simple in terms of trying to understand whether there are any differences between completers vs non-completers on a range of data that you have available, testing different assumptions of missingness, or conducting some form of imputation to “fill in the blanks”. This chapter provides overviews, guides, and software examples to conduct a range of missing data analyses.

8.1 Journal Articles and Textbooks

Van Buuren, S. (2018) Flexible imputation of missing data. Second edition. Chapman & Hall/CRC.

- *Excellent online textbook available here <https://stefvanbuuren.name/fimd/> written by the same author who wrote the mice package for R.*

Enders, C. K. (2022). Applied missing data analysis. Second edition. Guilford Press.

- *Probably one of the go-to textbooks in terms of missing data analysis. They have also developed their own program to conduct missing data analysis in complex samples/analyses that are currently not available in other programs (see Blimp in software section). <https://www.appliedmissingdata.com/>*

Sterne, J. A., Carlin, J. B., Royston, P., & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. <https://www.bmj.com/content/338/bmj.b2393>

- *Introduction and provides workflows and reporting guidelines for multiple imputation. It is commonly found that multiple imputation is a robust and widely used approach when data are missing at random.*

Allison, P. D. (2012). Handling missing data by maximum likelihood. SAS Global Forum. Paper 312-2012.

‘ *Comprehensive paper comparing multiple imputation to maximum likelihood when handling missing data and provides some examples when using SAS.* <https://statisticalhorizons.com/wp-content/uploads/MissingDataByML.pdf>

Baker, S. G. (2020). Maximum likelihood estimation with missing outcomes: from simplicity to complexity. *Stat Med*, 38(22), 4453-4474. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6879193/>

‘ *Paper outlining the use of maximum likelihood for clinical or prevention studies when you have partial missing data on the outcome. Starts to become more complicated as they outline different type of missingness analyses using maximum likelihood.*

Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. *BMC Medical Research Methodology*, 17, 162. <https://doi.org/10.1186/s12874-017-0442-1>

- *Pretty self-explanatory paper providing some good advice regarding the use of multiple imputation and other methods and when you should use them. But take some of these rules of thumb with caution.*

Grund, S., Ludtke, O., Robitzsh, A. (2016). Multiple imputation of multilevel missing data: an introduction to the R package pan. *SAGE Open*. <https://journals.sagepub.com/doi/pdf/10.1177/2158244016668220>

- *Discusses the additional complexity of using multiple imputation for multilevel data (e.g., when there are multiple levels or clustering in the data). Provides an introduction to the pan R package to address this issue.*

8.2 Websites

<https://www.publichealth.columbia.edu/research/population-health-methods/missing-data-and-multiple-imputation>

- *Provides a nice basic overview of missing data procedures, the different types of missing data, and potential options for handling missing data. Also provides an excellent list of resources in terms of textbooks, etc and does a better job at that than this list!*

https://stats.oarc.ucla.edu/stata/seminars/mi_in_stata_pt1_new/

- *Provides a worked example of how to conduct multiple imputation using Stata.*

https://stats.oarc.ucla.edu/sas/seminars/multiple-imputation-in-sas/mi_new_1/

- *Provides a worked example of how to conduct multiple imputation using SAS.*

<https://imaging.mrc-cbu.cam.ac.uk/statswiki/FAQ/missing>

- *A blog-style post from University of Cambridge regarding missing data in SPSS.*

<https://missingdata.org/why-multiple-imputation-is-better-than-maximum-likelihood/>

- *A blog post examining the benefits of multiple imputation over maximum likelihood to handle missing data. Also there is a follow-up blog on the same website about the number of imputations required.*

8.3 Online courses and videos

<https://vimeo.com/147353266> <https://vimeo.com/147723900>

- *A 6-hour online workshop conducted by Craig Enders on dealing with missing data*

<https://www.youtube.com/watch?v=vvrMc9cUrS0>

- *A walkthrough of multiple imputation using Blimp. More videos on using Blimp are available from here <https://www.youtube.com/@BlimpVideoSeries-rb1sz/videos>*

<https://www.youtube.com/watch?v=gdKTJdrqtOI>

- *A short course on missing data analysis via Mplus conducted as part of Johns Hopkins lecture series.*

https://www.youtube.com/watch?v=BS_bXe1dd7U

- *An introduction lecture on missing data handling with Mplus by Dr James Peugh at University of Kentucky.*

8.4 Software

BLIMP <https://www.appliedmissingdata.com/blimp>

- *A dedicated software package for conducting latent variable modelling and imputation for incomplete data with up to three levels (i.e. multilevel models). This program can do imputation when you have multilevel models such as time nested in individuals and individuals nested in schools, etc.*

Chapter 9

Logistic regression

This chapter provides resources on a specific type of regression that we often utilise at the Matilda Centre and one that is relevant to the field of mental health and substance use research and that is logistic regression. This type of regression differs from regular “linear” regression given that the outcome variable is considered categorical (either binary or multi-level) as opposed to continuous. So, for example, if you have a diagnostic variable that is your primary outcome coded as either 1 or 0 depending on whether the diagnosis is present or absent, then you could use logistic regression to determine significant associations with that diagnosis (in either a univariable or multivariable format). Standard logistic regressions tend to deal with binary outcomes, however it is possible to analyse categorical outcomes with more levels using either ordinal logistic regression (where there is a natural order to the categories) or multinomial logistic regression (where there is no order to the categories). This chapter will provide information on all regressions that involve categorical variables.

9.1 Journal Articles and Textbooks

Hosmer, D. W. and Lemeshow, S. (2013) Applied logistic regression. Wiley.

- *This book provides a great introduction to logistic regression and it is particularly helpful if you have already been introduced to linear regression. Chapters in the book cover topics such as multiple logistic regression, interpreting logistic regression results, model-building, assessing model fit, multinomial and ordinal logistic regression, and correlated binary data. A copy of this textbook is available free online via Usyd library https://sydney.primo.exlibrisgroup.com/permalink/61USYD_INST/12rahnq/alma991019090819705106*

Agresti, A. (2013) Categorical Data Analysis. Chichester: Wiley.

- *This text provide additional information on topics related to categorical data analysis over and above logistic regression, for example multinomial responses, loglinear models, random effects, etc. A copy of this textbook is available free online via usyd library https://sydney.primo.exlibrisgroup.com/permalink/61USYD_INST/12rahnq/alma991020517119705106*

Burkner, P-C. and Vuorre, M. (2019) Ordinal regression models in psychology: A tutorial. Advances in Methods and Practices in Psychological Science, 2: 77-101.

- *I guess the title of this article says it all. This article goes in detail for the analysis of ordinal variables in comparison to treating them as linear. It also provides some examples of how to run these analyses in R.*

Williams, R. (2014) Understanding and interpreting generalised ordered logit models. The Journal of Mathematical Sociology, 40: 7-20.

- *This article focuses on ordinal logistic regression and provides examples and specific models for analysing this type of data.*

Williams, R. and Jorgensen, A. (2023) Comparing logit and probit coefficients between nested models. Social Science Research, 109.

- *This article covers a niche but important topic in logistic regression, that is the problem of comparing coefficients or odds ratios between nested logistic regression models (a common practice in linear regression). This article discusses the problem, why we should be careful, and solutions for when you do want to compared coefficients across nested models. Highly recommended that you read this article*

VanderWeele, T. J. and Knol, M. J. (2014) A tutorial on interaction. Epidemiological Methods, 3: 33-72.

- *Another niche but highly important article focusing on interpretation of interactions and differences between additive and multiplicative scales in relation to linear, log-linear, and logistic regression models. Again highly recommended reading*

Chen, Q., Nian, H., Zhu, Y. et al. (2016) Too many covariates and too few cases? - A comparative study. Statistics in Medicine, 10: 4546-4558.

- *An article on the specific topic of what to do when you want to look at many covariates but have small sample sizes in logistic regression. An interesting article but only worth a read if you find yourself in this situation.*

Johfre, S. S. and Freese, J. (2021) Reconsidering the reference category. *Sociological methodology*, 51: 253-269.

- *This article covers an important topic in logistic regression, the choice of reference categories. This mostly reflects the choice of reference categories when using categorical explanatory (or independent), so not strictly applied to only logistic regression, but something that we do often.*