Dyadic Structural Equation Modeling in R

John Kitchener Sakaluk

2024-10-15

Table of contents

4

Preface

This is a book about analyzing dyadic data with latent variables using the structural equation modeling (SEM) framework.

Academics often find reading their own writing to be a "cringe-worthy" experience, and reading that leading sentence is a particularly strange experience for me. I never once thought I'd write a book on something so statistics- and programming-oriented like dyadic SEM:

I dropped out of 11th grade programming course

I nearly failed my 12th grade introductory statistics course

My formal training in dyadic data analysis consists entirely of one weeklong workshop.

And yet...

This is a book about analyzing dyadic data with latent variables using the SEM framework, and I am writing it. And I hope it will prove to be a book worth reading, if you are interested in dyadic data analysis. Why? Because despite my aforementioned reasons to feel surprised that I am writing this book, I *do* know a thing or two about SEM. And maybe in spite of my earlier challenges with statistics and programming–actually, perhaps *because* of them–I will be able to provide you a different way of thinking about dyadic data analysis. These challenges have shaped the way I teach SEM-related content, and directly informed many of the design choices for how I have programmed the *R* package *dySEM*

Indeed, cast through the lens of SEM, the analysis of dyads affords unique opportunities to ask–and answer–some particularly interesting questions. Moreover, dyadic SEM is used so infrequently, as of the writing of this initial draft, so that I think the "market" to make an impact with this approach to data analysis is *wide open*. I hope, in time, through writing this book and developing open-source software like *dySEM*, that this "market" will become a bit more lively.

The good news, if I have your interest, is that I plan to write and maintain this book open-access. That means it will be legally free for you to use whenever you want. **The bad news** is that I plan to write and maintain this book open-access. That means that I will write it in bursts–not always in a straightforward linear way–and things will change. Don't be surprised to come back between versions and find I've totally reorganized a section, or a chapter, or even an "Act" of the book. And of course, all of my embarrassing mistakes and gaps in knowledge will be on display for all to see.

With that in mind, if you notice conceptual or code-based errors, or have requests for features or area of content converage, you'd be doing me a favour by submitting these to the "Issues" page (LINK) of the GitHub repo, where I'll be maintaining the book.

Otherwise, I hope you enjoy this book and find somethign useful in it.

Part I

Act I: Getting Started

1 Introduction

This is a book about analyzing dyadic data with latent variables using the SEM framework. As such, it occupies a unique location within the market of available books across several domains. Kenny, Kashy, and Cook (2006), for example, is an excellent book focused exclusively on dyadic data analysis. It describes the challenges of managing dyadic data, as well as many of the most popular models of dyadic data. But though Kenny, Kashy, and Cook (2006) describe the SEM framework and its application to some models, they don't really mean "SEM" in the way that I do. That is, their consideration of SEM isn't particularly engaged with the prospect of modeling latent variables, while I am *exclusively* focused on the use of latent variables.

A book like Little (2013), meanwhile, describes the use of latent variable modeling for dependent data *like* dyadic data, albeit in the context of longitudinal research designs, where the source of dependency is repeated observation. As such, though this book possesses some wisdom for those wishing to use latent variables in their models of dyadic data, the book is not written with dyadic data or dyadic models in mind. And as it turns out, things get *weird* when you are analyzing dependent groups of two (as opposed to, say, individuals over the span of three, four, or five waves of repeated assessment).

There are many other excellent books, too, that offer additional guidance–to some (in)direct degree or another–for applying latent variable models to dyadic data, such as Brown (2015) and Kline (2023) and Bolger and Laurenceau (2013). By writing this book, I mean to take nothing away from their value propositions; they are great resources. But they either do not consider the unique features of dyadic data and/or do not substantively engage in the distinctive benefits to modeling dyadic data with latent variables.

This is a book about analyzing dyadic data with latent variables using the SEM framework. And so I will exclusively discuss the analysis of dyadic data, **and** I will exclusively discuss its analysis with latent variable models.

In **Act I** of the book, I will first lay out "The Big Picture" (Chapter 2) of the what's, why's, and how's of dyadic SEM with latent variables, as well as introducing some of the technical jargon I will use throughout the rest of the book. I then discuss some of the unique considerations of data management for dyadic data analysis (Chapter 3), before providing an overview of latent variable theory in the context of modeling dyadic data (Chapter 4).

In **Act II** of the book, I attempt to provide what I consider to be a "sufficient" overview of the conceptual and applied specifics of modeling latent variables, without yet engaging with how to extend this framework to the analysis of dyadic data. Chapter 5 is essential reading

if you are unaccustomed to the statistical features of latent variables models (e.g., the visual depictions, notation and interpretation for particular features, etc.,) and common analytic practices within them (e.g., fixing or constraining parameters). I then discuss two related problems that must be resolved in fitting SEMs (model identification and setting the scale of the latent variable(s)) (Chapter 6), before they can be estimated (Chapter 7). These topics then set the table for describing how we evaluate (Chapter 8) and compare (Chapter 9) SEMs. I then work through all of this (and more) in an applied non-dyadic example (Chapter 10).

I then pause in **Intermission 1** to provide some foreshadowing about the dyadic SEMs we will encounter in the subsequent chapters (Chapter 11). For those approaching this book while having some familiarity of the models described in Kenny, Kashy, and Cook (2006), this chapter will help to transition you to thinking about these models recast in latent space.

In **Act III**, we *finally* get into the specification of models for dyadic data with latent variables, beginning with the simplest cross-sectional models possible: those intended to capture only one *construct* (i.e., "uni-construct") shared somehow between dyad members. These include the dyadic one-factor model (Chapter 12), the correlated two-factor model (Chapter 13), the bifactor model (Chapter 14), and the hierarchical factor model (Chapter 15). These models have a surprisingly interesting (and complex) relationship to one another, which I discuss in the subsequent chapter (Chapter 16). I also describe how to use dyadic invariance testing within these models, in order to evaluate the generalizability of latent variable model parameters across partners (Chapter 17), which plays an important role in many other comparisons in dyadic data analysis, as well as being an (unappreciated) interesting phenomenon in its own right. I conclude this section with a discussion of an important but vastly underappreciated issue: how to choose among competing uni-construct models for a given set of dyadic data (Chapter 18).

In **Act IV**, we move to discussing dyadic SEMs that are latent embodiments of the kinds of models that may seem more prototypical in dyadic data analysis (i.e., those covered in Kenny, Kashy, and Cook (2006)). These models involve the prediction of one dyad-related construct by another (i.e., are bi-construct). I first describe bi-construct models where the predictor construct and outcome construct share the same uni-construct dyadic model, including the Couple Interdependence Model (Chapter 19), the Actor-Partner Interdependence Model (Chapter 20), the Bifactor Structural Model (Chapter 21), and the Common Fate Model (Chapter 22). I conclude this section with a discussion (and some encouragement) of how different uni-construct models could be combined in more boutique bi-construct models (Chapter 23).

We then pause once more, in **Intermission 2**, in order to discuss–with the knowledge of uni-construct and bi-construct dyadic SEMs under our belts–just how complicated the concept of "distinguishability" is, when cast through the SEM lens.

Finally, in **Act V**, we delve into even more complex applications of dyadic SEM, including some themes of practice that are not yet done, yet which I hope will be on the (near) horizon of analytic practice in our field. These include the modeling of so-called "third variable" processes (Chapter 25), testing dyadic SEMs across groups (Chapter 26), the modeling of both dyadic

and longitudinal dependency with latent variables (Chapter 27), and the deployment of data-driven exploratory models to provide a plausible dyadic measurement model (Chapter 28). I also discuss the application of (and need for more) Monte Carlo simulation studies (Chapter 29), to evaluate the performance of dyadic SEMs (and other modeling strategies); in this chapter, I also discuss how these simulations can help to inform sample size planning. I then end this section with some encouragement and guidance of how to contribute to open-source dyadic data modeling tools (Chapter 30), for those so inclined.

I'll also (eventually) write a Conclusion to this book (Chapter 31), and I'm sure it'll be very meaningful and impressive. But for now, I need to generate some content, before I can realize what it is I ought to conclude.

What I will *not* write about, however, is the basics of using R–the open-source cross-platform statistical programming language that I use in my dyadic SEM work (and upon which this book and its applications currently rely). If you are entirely new to R, the good news is that most dyadi SEM modeling instantiations require precious little of fiddling around with basic data management in R. That is, you're often "good to go" soon after data importation. If you need additional scafolding for using R, however, I encourage you to check out "R for Data Science", or "R4DS" as it's sometimes known (**NEED REFERENCE**). It's a gold mine of useful information for R users of all levels of comfort.

And no: I will not provide analytic resources for other programming/statistics languages (e.g., SPSS/AMOS, SAS, MPlus). Though I have sometimes done this in the past (e.g., John K. Sakaluk and Short (2017), John Kitchener Sakaluk (2019)), and some of the models I describe herein are possible to specify in these other softwares, I have decided that I am done supporting proprietary software. This may lose me some readers; so be it. I want to create learning resources and tools that are avaiable to anyone and everyone, for free, and the increasing expense of these other software packages threatens what I see as a necessary mandate to democratize access to learning.

2 The Big Picture

This is a book about analyzing dyadic data with latent variables using the SEM framework, and I think your experience reading the remainder of the book will be aided by providing you with a rapid-fire tour of the bigger themes within it (and an introduction to some key jargon).

So, let us begin with that introductory phrase: analyzing *dyadic data* with *latent variables* using *the SEM framework*.

2.1 What is "dyadic data"?

Putting a finger on what dyadic data is, specifically, is surprisingly difficult when you begin to think of variety of ways in which the organisms we study (people, animals, nation-states, etc.) can, and do, "pair up". And so in some ways, it's easier to begin by stating what dyadic data *is not*.

Dyadic data is not data from organisms that have nothing in common. How much nothing? Total. Imagine you ran a goofy study where you experimentally assigned people to pairs and estimated the correlation between pairs on *Trait X*. In the long run, it should be $r = .00$, right? Okay: those data are *not dyadic*. You should be fine to structure these data in such a way that allows you to analyze them as coming from independent observations, and go on with your business.

Eliminating data from organisms that have nothing in common still leaves a lot of possible contexts on the table as potentially dyadic. At the opposite end of the spectrum of *dyadic-ness* are data that come from organisms that totally overlap. Though contexts like this are rare (rare enough that I can only think of: data from a sample of identical twins), they probably would satisfy our intuition-based criteria of what is dyadic data.

The tricky cases are those in-between: are data from research where organsism have something (but not everything) in common, *dyadic* data? For example, what if you collect data from a classroom ($n = 200$) and a couple of pairs of individuals (*pair 1* and *pair 2*, who otherwise do not interact) happen to share a hometown? Or in a sample of beetles ($n = 60$) a few pairs (*pair 1*, *pair 2*, and *pair 3*), happen to be siblings from the same brood?

Though examples like these contain *dyads*, I would not call the affiliated datasets *dyadic*. This is because the paired nature of some observations is inconsistent; there are a few dyads/pairs,
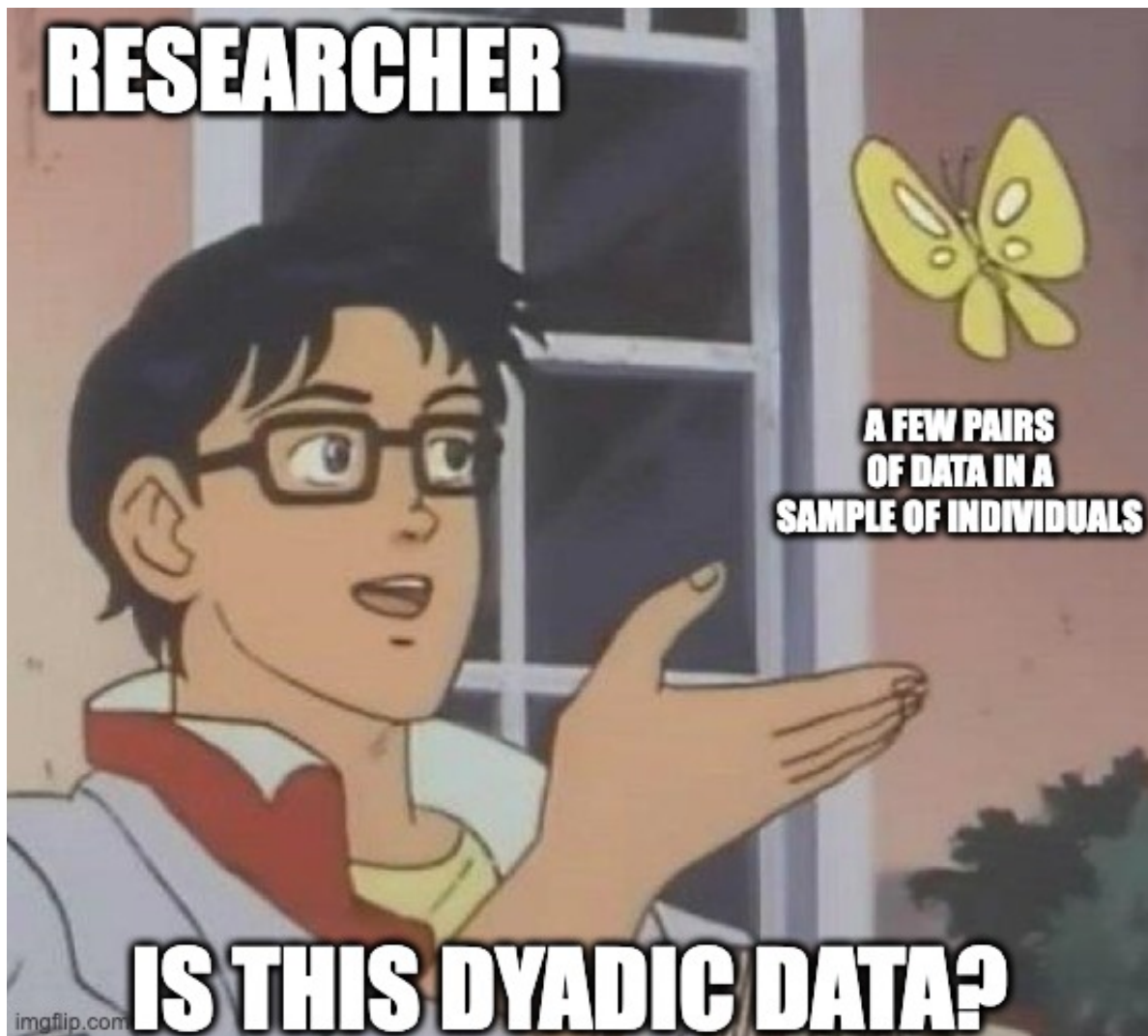
Figure 2.1: The first of many stats-related memes

but most of the data otherwise comes from individuals, and it's not clear how a researcher would systematically be able to identify the linked observations between the small number of dyads.

Instead, **_let us consider data to be "dyadic data" if it is dyadic by design_**. That is, a researcher voluntarily deployed an approach to collecting data in such a way that all organisms in the sample are members of a dyad, known to the researcher. In this book, we will exclusively deal with an approach to analysis of quantitative dyadic data, though qualitative dyadic research designs are also possible (at least for human-based samples).

We will see in Chapter 3 that there are further ways we can conceptualize and categorize dyadic data–and the analytic techinques in this book are most useful to only one of these dyadic data designs–but that is a sufficient understanding of "dyadic data" for now.

2.2 What are "latent variables"?

What, then, are these so-called *latent variables*, and what is useful or interesting about them? I like Dictionary.com's definition of *latent* as a starting place:

> [leyt-nt] present but not visible, apparent, or activated; existing as potential

But what use is a variable that you cannot directly observe? Social scientists have been interested in latent variables since as far back as the late 1800's and early 1900's. A classic example is Spearman (1904) (of the "Spearman correlation coefficient" fame)(REF). Spearman was psychologist who was interested in, among other things, mental abilities. At the time, there were a number of competing theories of ways to categorize "types" or "kinds" of mental abilities (e.g., visual, memory-based, spatial, etc.,), and how to organize them. Spearman's contributition to this literature was noteworthy, as he provided evidence for a *latent* kind of *generalized intelligence* (which he dubbed "g")–not a form of mental ability you could directly see or appraise, but which shaped performance on all tests of various mental abilities to one extent or another. The impact of this work was tremendous, both in terms of its theoretical value and the way it contributed to subsequent statistical methodology. With respect to the former, Spearman helped to establish the plausibility of latent variables (of which generalized intelligence may just be one) as important determinants of social and psychological processes. Methodologically, meanwhile, through is work, Spearman "casualy" provided one of the first exemplars of factor analysis (and a very specific kind of it, at that)–an analytic approach that would come to play a central role in a staggering amount of future research. Indeed, factor analysis is the beating heart of many of the models of dyadic data we will consider in this book.

The notion of latent generalized intelligence may seem reasonably intuitive, but what of latent variables in dyadic data? Here, I must profess, it's easy for me to provide exemplars that

work in the discipline of psychology, and much more difficult to provide examples that work in disciplines farther afield (e.g., ecology). Still, let me try.

For my money, the game-changing moment in Psychology to open the door to the study of latent variables was the Cognitive Revolution. The theoretical coin of the realm before the Cognitive Revolution was, for many decades, Behaviorism. And under Behaviorism–specifically, the edicts of "Methodological Behaviorism" (WATSON 1924 REF)–the experience of thought and feeling was thought to be bereft of scientific value. Only that which could be seen, directly–behavior–was amenable to scientific inqiury. The Cognitive Revolution was largely responsible for upending this ban on studying thoughts and feelings, and with that, it became open-season to scientifically study intrapsychic phenomena like attitudes, beliefs, emotions, motives, and values. These concepts represent the kind of psychological terrain where wild latent variables may be found, and studied.

2.3 What is "structural equation modeling (SEM)" ?

2.4 Bringing it all together

2.4.1 What is "dyadic structural equation modeling"?

2.4.2 Why use dyadic SEM vs. Alternative Modeling Frameworks?

2.5 Uni-construct vs. Bi-construct (and Beyond) dyadic SEMs

2.6 Fitting Dyadic SEMs

2.6.1 Software Choices

2.6.2 Using the *dySEM* and *lavaan* Packages for *R*

2.7 Pushing the Boundaries of dyadic SEM

3 Dyadic Data Structure

4 Latent Variables in SEM

Part II

Act II: SEM Fundamentals

5 Models (and Their Elements)

6 Model Identification and Scale-Setting

7 Estimating Dyadic SEMs: Common Practices, Problems, and Solutions

8 Model Fit Indexes

9 Model Comparison Strategies

10 A Simple (Non-Dyadic) Example in {lavaan}

Part III

Intermission 1: Meet Your Dyadic Models

11 Design, Scope, and Form of Dyadic Models

Part IV

Act III: Uni-Construct Cross-Sectional Dyadic Models

12 The One-Factor (Surprinslgy Dyadic) Model

13 The Correlated Two-Factor Model

14 The Bifactor Model

15 The Hierarchical Factor Model

16 The Causal Relations Between Univariate Cross-Sectional Dyadic SEMs

17 Dyadic Invariance Testing in Univariate Cross-Sectional Dyadic SEMs

18 Choosing Between Uni-Construct Cross-Sectional Dyadic Models

Part V

Act IV: Bi-Construct Cross-Sectional Dyadic Models

19 The Couple Interdependence Model

20 The Actor-Partner Interdependence Model

21 The Bifactor Structural Model

22 The Common Fate Model

23 Mixing Univariate Structures in Bivariate Models

Part VI

Intermission 2: What Is '(In)distinguishability'?

24 (In)distinguishability Is More Complicated than You Think

Here is an example of an equation where different parts are color-coded:

$$E = mc^2 \quad \text{where} \quad E = m \cdot c^2$$

Part VII

Act V: Advanced Applications of Dyadic SEM

25 "Third Variable" Cross-Sectional Dyadic Models

26 Multi-Group Cross-Sectional Dyadic Models

27 Longitudinal Dyadic Models

28 Dyadic Exploratory Factor Analysis

29 Monte Carlo Simulations of Dyadic Modeling Strategies

30 Contributing to Open-Source Software Development for Dyadic SEM (and Beyond)

Part VIII

Fin

31 Conclusion

References

Bolger, Niall, and Jean-Philippe Laurenceau. 2013. *Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research.* New York, NY: Guilford Press.

Brown, Timothy A. 2015. *Confirmatory Factor Analysis for Applied Research.* 2nd Ed. New York, NY: Guilford Press.

Kenny, David A., Deborah A. Kashy, and William L. Cook. 2006. *Dyadic Data Analysis.* New York, NY: Guilford Press.

Kline, Rex B. 2023. *Principles and Practice of Structural Equation Modeling.* 5th Ed. New York, NY: Guilford Press.

Little, Todd D. 2013. *Longitudinal Structural Equation Modeling.* New York, NY: Guilford Press.

Sakaluk, John Kitchener. 2019. "Expanding Statistical Frontiers in Sexual Science: Taxometric, Invariance, and Equivalence Testing." *The Journal of Sex Research* 56 (4-5): 475–510. https://doi.org/10.1080/00224499.2019.1568377.

Sakaluk, John K., and Stephen D. Short. 2017. "A Methodological Review of Exploratory Factor Analysis in Sexuality Research: Used Practices, Best Practices, and Data Analysis Resources." *The Journal of Sex Research* 54 (1): 1–9. https://doi.org/10.1080/00224499.2015.1137538.

Spearman, C. 1904. ""General Intelligence," Objectively Determined and Measured." *The American Journal of Psychology* 15 (2): 201–92. https://doi.org/10.2307/1412107.