

Multilevel Thinking

Using Multilevel Models To Thread The Needle Between Universality and Cultural Specificity in Cross Cultural Research

Andrew Grogan-Kaylor

12/22/22

Table of contents

1. The Usefulness of Multilevel Modeling and Multilevel Thinking.....	2
Acknowledgements.....	3
Some Preliminary Thoughts	4
2. Introduction	5
2.1 Quantitative Methods and Social Justice.....	5
2.2 Are Answers from Social Science “Obvious”?	6
2.3 Presenting Advanced Statistical Ideas	7
2.4 Research on Parenting and Child Development in International Context.....	7
2.5 Universalism And Particularity	8
2.6 Multilevel Models As The Study Of Variation.....	9
3. Simulated Multi-Country (Multilevel) Data	13
4. Software.....	15
5. Conceptual Framework.....	15
5.1 Units of Analysis	15
5.2 Variables and Processes at Multiple Levels	16
6. The Cross Sectional Multilevel Model.....	16
6.1 The Equation	16
6.1.1 Two Levels of Equations.....	17
6.1.2 One Level of Equations.....	17
6.2 Estimating Standard Errors And p Values	18
6.2.1 Introduction	18
6.2.2 Compare OLS and MLM.....	19
6.3 Multilevel Structure	19
6.3.1 Use Some Data Simulated For This Particular Example	20
6.3.2 Graphs	20
6.3.3 Regressions	21

6.3.4 A Thought Experiment	22
6.4 Regression With Simulated Multi-Country Data	22
6.5 Summary of Advantages Of The Multilevel Model.....	23
6.6 Predicted Values.....	24
6.7 Variation.....	24
6.7.1 Measured and Unmeasured Variation	24
6.7.2 Unconditional Model	24
6.7.3 Intra-Class Correlation Coefficient.....	24
6.7.4 Variation In Intercepts or Outcomes	26
6.7.5 Variation In Predictors.....	26
6.7.6 Variation in Slopes	26
6.7.7 Variation As An Outcome.....	26
6.7.8 Maximal Models	27
6.8 Some Wrong (or Partially Wrong) Approaches.....	27
7. The Longitudinal Multilevel Model	28
7.1 Use Data With Multiple Observations Per Individual	28
7.2 The Equation	30
7.3 Regression With Simulated Multi-Country Longitudinal Data	30
7.4 Autocorrelation.....	31
7.5 Causality	31
7.5.1 The Importance of Causal Reasoning	31
7.5.2 Formal Criteria of Causality	32
7.5.3 Simpson's Paradox	33
7.5.4 A Simpler Multilevel Model To Explore Causality	34
7.5.5 Fixed Effects Regression	34
7.5.6 The Correlated Random Effects Model.....	35
8. Conclusion	36
References	36

1. The Usefulness of Multilevel Modeling and Multilevel Thinking

For decades now, multilevel models have been an important quantitative tool for social research. While multilevel models have become ubiquitous in social research, there are dimensions of these models that are explored less frequently in published articles. This document arises from my experiences of teaching a course entitled *Multilevel and Longitudinal*

Modeling that I have taught for over a decade in the *Joint Doctoral Program in Social Work and Social Science* at the University of Michigan.

My contention is that *multilevel modeling* offers powerful tools for understanding the *multilevel data* that social researchers often confront. For example, researchers are often interested in studying outcomes for diverse groups of children in different schools, residents of diverse and different neighborhoods, or individuals or families living in diverse and different countries. Such inherently multilevel data lead to analytic complexities, some of which appear to me to be well understood, while others seem to be much less often appreciated.

The point that I wish to make about multilevel data is that when presented with complex multilevel data, failure to use the appropriate multilevel model may lead to conclusions that are demonstrably incorrect. Fortunately, many of these difficulties can be avoided with applications of simple and straightforward multilevel models.

After presenting some initial ideas about multilevel modeling, I go on to explore some more complex ideas about multilevel models that I see less often in the published empirical literature. I focus especially on the idea of *multilevel models as the exploration of variation across countries and cultures*.

Certainly, none of the statistical ideas contained in this document are unique to me. There are thorough—and often much more mathematically rigorous—presentations of many of the ideas contained in this document in some of the excellent foundational texts on multilevel modeling such as the early book by Raudenbush & Bryk (2002), the excellent book on longitudinal models by Singer & Willett (2003), and Rabe-Hesketh & Skrondal (2012)'s more recent and extremely comprehensive two volume text. Luke (2004), and Kreft & de Leeuw (1998), offer shorter, less mathematical, but still excellent introductions to the topic of multilevel modeling. Gelman et al. (2007) introduced me to the ideas that in this document I describe as “multilevel structure” using an example with voting patterns.

My intent in this document is to offer a kind of accessible tutorial for applied researchers, including especially those who see their research having some advocacy based component. My approach, while offering up some equations, is less mathematical than some of the above mentioned texts, and written with the intent of providing a clear and practically focused guide for the applied researcher who is attempting to carry out better research with diverse populations.

Acknowledgements

No good learning happens without community. At least, that has always been true for me. I am grateful for many creative and energizing discussions with the other members of the MICS (UNICEF data) research team: Professor Shawna Lee, Professor Julie Ma, Dr. Kaitlin Ward, and Professor Garrett Pace. I'm thankful for their collegiality, their friendship, and their dedication to good science. I'm also very grateful to one of my mentors, Professor Sandra Danziger, who has taught me so much about the *what* and the *why* of mentoring, teaching, and doing research. I'd like to thank Ross Grogan-Kaylor for continued interest in the progress of this document, and probing thoughtful questions. Shari Grogan-Kaylor had many important and challenging questions about the *what* and the *why* of this document. Don Deutsch showed ongoing interest in the development of this document, and has asked some hard questions that have improved its logic. Importantly, I'd like to express gratitude to the many students in my class on *Multilevel and Longitudinal Modeling* who over the years have helped me think more deeply about statistical and substantive issues, including Dr. Kaitlin Ward, Professor Garrett

Pace, Professor Julie Ma, Professor Berenice Castillo, Professor Maria Galano, Dr. Sara Stein, Madhur Singh, and Tong Suo. Lastly, I'd like to thank a few people I've ever met: Pau Dones for his music and inspiration, and Mary Oliver, David Whyte and John O'Donohue for their wise words. While I'm thankful for the inspiration and collegueship provided by others, any remaining errors and omissions in this document are of course my responsibility.

Some Preliminary Thoughts

Preliminaries

"Like you I

Love love, life, the sweet smell of things, the sky-blue landscape of January days.

...

I believe the world is beautiful.

And that poetry like bread, is for everyone.

And that my veins don't end in me.

But in the unanimous blood.

Of those who struggle for life,

Love, little things,

Landscape and bread, the poetry of everyone."

— Roque Dalton (tr. By Jack Hirschman)

"A lifetime is too narrow to understand it all, beginning with the huge rockshelves that underlie all that life.

No one ever told us we had to study our lives, make of our lives a study, as if learning natural history or music, that we should begin with the simple exercises first and slowly go on trying the hard ones, practicing till strength and accuracy became one with the daring ...

But there come times—perhaps this is one of them—when we have to take ourselves more seriously or die, when we have to pull back from the incantations, rhythms we've moved to thoughtlessly, and disenthral ourselves, bestow ourselves to silence, or a severer listening ..."

— Adrienne Rich

"Research is formalized curiosity. It is poking and prying with a purpose."

— Zora Neale Hurston

2. Introduction

2.1 Quantitative Methods and Social Justice

There is clearly need for both qualitative and quantitative methods. Central to the argument of this document is the idea that advanced quantitative methods can be core contributors to the agenda of understanding issues of diversity and social justice more fully and thoroughly (Cokley & Awad, 2013; Grogan-Kaylor et al., 2018). Quantitative methods, particularly in discussions comparing qualitative and quantitative methodologies, are sometimes labelled as inherently *positivist* methods. My argument regarding this point is twofold. There is nothing within the mathematics of quantitative methods that requires a positivist epistemology. Quantitative methodologies could as easily be conducted using a critical epistemology—that is aware of dynamics of power and privilege—as any other methodology (Stage & Wells, 2014). Second, when we have samples of a hundred, several hundred, several thousand, or even hundreds of thousands of study participants, it is difficult to imagine a methodology other than a quantitative methodology that could accomplish the following:

1. Sift through thousands of responses, and determine the *overall, or average, pattern of relationships* between risk factors, protective factors, and outcomes.
2. Explore the *variation in these relationships* across social contexts.
3. Determine whether there is evidence that the relationships observed within the data are more than *statistical noise*.
4. Adjudicate the *complex multivariate relationships* of risk factors, protective factors and outcomes.

In [Section 6.2](#) and [Section 6.3](#), I explore the ways that *multilevel* data can contribute substantially to the complexity of this enterprise. I thus argue that quantitative methods can play an important role in contributing to liberatory ideas. I note that one of the pioneers of liberation psychology, Martin-Baró (Aron & Corne, 1994), used both qualitative and quantitative methods (Martin-Baro, 1994), including in the latter case, relatively sophisticated arguments about patterns of missing data across a survey data set (Aron & Corne, 1994).

There is thus an ethical argument that is embedded in this document. Many of us do research with the hope of better understanding the relationship of risk and protective factors with outcomes in diverse, and often disadvantaged or marginalized, populations. Many of us further hope that our work might be part of conversations about appropriate policies, programs, treatments or interventions. Given the frequent vulnerability and marginalization of the people with whom we work, when using quantitative methods, it is incumbent upon us to employ methods that adequately address the complexities of the data, that offer an appreciation of the variability and diversity within the data, that provide the most accurate estimates possible, and that increase the probability of obtaining correct answers to important substantive questions.

“It is hard to imagine that anyone with a humanitarian worldview would argue against the need for a more quantitatively literate citizenry. Informed political decision-making, retirement planning, active parenting, and the vast majority of choices we make in our personal, occupational, and civic lives can be better served by improved quantitative understanding and reasoning, as well as accompanying action-oriented dispositions.” (Wiest et al., 2007)

The idea of this document is that a deeper study of multilevel modeling can result in an advanced “quantitative literacy” (Wiest et al., 2007), or “principled argument” (Abelson, 1995), that is appropriate for drawing accurate conclusions from multilevel data.

2.2 Are Answers from Social Science “Obvious”?

Closely related, I think to the the idea that quantitative research can advance issues of social justice, is the question of whether answers from social science are “obvious”. If social science answers are obvious, then social science has limited abilities to make new discoveries, and to build scientific foundations for evidence.

I have been thinking a lot about the idea that *Everything Is Obvious, Once You Know The Answer*, as detailed in the book with this title by Duncan Watts (2011).

This seems to me especially true in social research. Arguably, some conclusions of social research may indeed be obvious. For example, it may be obvious that *Adverse Childhood Experiences* (ACEs) are associated with long term decreases in mental health. However, even obvious conclusions may need to be quantitatively documented, in order to legitimate programs and interventions, and to secure funding. I also observe that I think that there is often a *historical* dimension to what is considered “obvious”: conclusions that are at first considered to be unlikely to be true, or even counter-intuitive, require the weight of accumulating evidence over time for these connections to become “obvious”. It is likely that the “obviousness” of the relationship between ACEs and later physical and mental health problems did not become apparent until research began to document these relationships (e.g. Felitti et al. (1998)). As another example, Proctor (2012) documents the way which smoking was first considered to be an *unlikely* cause of lung cancer; only over the course of several decades of research and discussion to become an *obvious* cause of lung cancer. A similar *historical* dynamic seems to be playing out in some research on parenting and child development. Despite decades of evidence indicating that corporal punishment has undesirable consequences for children (Gershoff & Grogan-Kaylor, 2016b), corporal punishment remains a disciplinary strategy endorsed by the majority of the American population (Hines et al., 2022).

In contrast sometimes the conclusions of social research may not always be obvious. For example:

1. There has been an ongoing debate about whether corporal punishment is more or less harmful when used by parents in social contexts, or communities where it is more common, or normative. Eamon (2001) suggested that “when environmental risk is high, parenting practices that are firmer and higher in control result in lower levels of young adolescent antisocial behavior.” This echoes similar research by (Deater-Deckard et al., 1996) suggesting that physical punishment was harmful for European-American children, but not for African-American children. Later, larger sample research has found that this appears not to be the case: physical punishment is harmful for children in all groups (Gershoff & Grogan-Kaylor, 2016b, 2016a; Pace et al., 2019).
2. Using MICS Data (UNICEF, 2021), we conducted a study of the link between gender inequality and physical child abuse (Ma et al., 2022). We expected to find that higher levels of gender inequality led to higher levels of physical abuse for female children, but not for male children. Instead, we found that higher levels of gender inequality were associated with higher levels of physical abuse for both male and female children. Additionally, there was some slight evidence that male children were at higher risk of being abused than female children. Equally interesting was that we found that gender inequality was predictive of levels of child abuse, while country level GDP was not.
3. In a study of parenting during Covid-19 (Lee et al., 2022), we expected to find that households with children would experience *higher* levels of anxiety and depression than households without children. Instead, we found the opposite. Being in a household with children was generally *protective* against anxiety and depression.

In [Section 2.6](#), [Section 6.2](#) and [Section 6.3](#), I provide specific examples of how multilevel data provides even more opportunity to present answers that are *not* obvious.

2.3 Presenting Advanced Statistical Ideas

In presenting advanced, statistical concepts, one is faced with a quandary. One can present statistical concepts in the most general terms, in terms of x and y . While perhaps the mathematically most general way to present ideas, a highly general (and abstract) presentation risks not being a good way of teaching the ideas, as it is sometimes difficult to apply abstract ideas to one's own specific area of research.

Alternatively, one can present statistical ideas in terms of specific substantive concepts. The risk of making use of a specific substantive concept is that while concrete examples are always helpful, it may be difficult for the reader to generalize from a specific example to their own area of research.

I ground this presentation in research that we have conducted on parenting and child development in international context (Grogan-Kaylor et al., 2021; Ma et al., 2022; Pace et al., 2019; Ward, Grogan-Kaylor, Ma, et al., 2021; Ward, Grogan-Kaylor, Pace, et al., 2021; Ward et al., 2022). For the presentation in this document, I use simulated data on these issues.

Using the simulated data, I refer to *predictors* and *outcomes*, and explore the ways that the multilevel model can contribute to understanding how relationships between predictors and outcomes might be similar, or might be different, across *social contexts*. In the examples presented below, I focus on two predictors, parental *warmth*, and parental use of *physical punishment* and focus on the *outcome* of *improved* mental health. I use the social context of different *countries* in our example.

It is my belief that while I use this specific set of examples, that the idea of studying *families in different countries* is generalizable enough to a multiplicity of diverse contexts, such that the reader can apply these ideas to their own area of interest, whether that be *children in schools*; *residents in neighborhoods*; or *people in different countries*.

2.4 Research on Parenting and Child Development in International Context

Research on parenting and child development has identified robust associations between parenting behaviors and child developmental outcomes. Broadly speaking, physical punishment is associated with increases in child aggression, child anxiety and child mental health problems (Gershoff & Grogan-Kaylor, 2016b), while warm and supportive parenting is associated with decreases in these outcomes (Khaleque & Rohner, 2002; Rothenberg et al., 2022). However, much of this research is conducted on North American samples (Draper et al., 2022; Henrich et al., 2010).

Barth & Olsen (2020) have argued, that children constitute a class of oppressed persons. If children are oppressed, then it is imperative to empirically determine what factors are promotive of children's well-being, and what factors constitute risk factors that contribute to decreases in children's well-being. Equally imperative—given the North American focus of so much research on parenting and child development (Draper et al., 2022; Henrich et al., 2010)—would be efforts to extend the study of parenting and child development to a broader, more global context. As part of such a research agenda, it is necessary to have quantitative tools that are able to determine the consistency of relationships in parenting and child development. That is, are the relationships between certain forms of parenting and child developmental outcomes, largely consistent across countries, largely different across countries, or somewhere in between?

This document will discuss the ways in which a multilevel statistical perspective not only allows one to appropriately analyze cross cultural or international data, but also the ways in which a multilevel perspective affords the opportunity for more precise quantitative thinking about cross cultural phenomena.

This document takes a very pragmatic and very advocacy oriented approach to improving research on families and children, with the aim of improving the well-being of families and children.

“It shouldn’t be theories that define the problems of our situation, but rather the problems that demand, and so to speak, select, their own theorisation.” – Martin-Baro (1998) in Burton & Kagan (2005).

Following from this pragmatic and advocacy oriented emphasis, the document is largely oriented to the *doing* of quantitative social research with multilevel (or multi-country) data, and is therefore mostly statistical in nature.

The document moves quickly into detailed statistical arguments. Some of these statistical discussions may seem very technical, or even overly technical. However, an overarching theme of the document is that multilevel data contains hidden complexities. A lack of awareness of the complexities of multilevel data—e.g. complexities of multi-country data—might lead to statistical analyses that point in the wrong direction: yielding false positives; false negatives; or substantively wrong conclusions.

2.5 Universalism And Particularity

The specific domain of cross-cultural research on parenting and child development raises more general questions in cross-cultural research of *universalism* and *particularity*. With regard to child development it is universal that all children need some amount of emotional and material care to grow into healthy youth and healthy adults (Kottak, 2021). Further it is broadly understood that children should be protected from violence (UNICEF, 2014). This broad consensus is manifested in such documents as the Convention on the Rights of the Child (United Nations General Assembly, 1989) and the United Nations Sustainable Development Goals (United Nations, 2022), representing global efforts to ensure the children are cared for, and are protected against violence.

At the same time, broad international efforts to improve children’s well-being must engage with important considerations of cultural uniqueness. Put simply, what is considered to be beneficial for children in one country or culture may not be considered to be beneficial in all countries or cultures. Similarly, what is considered to be detrimental in one country or culture may not equally be considered to be detrimental in all. Within the area of parenting and child development, most of the debate has focused around the question of whether physical punishment is equally detrimental in all settings, particularly whether physical punishment is detrimental in countries where it is especially common, or normative (Gershoff et al., 2010). Much less attention has been focused on the study of positive parenting internationally, and the degree to which the outcomes of positive parenting are consistent across countries remains understudied (Ward, Grogan-Kaylor, Ma, et al., 2021).

However, as global initiatives to improve child well-being and family life move forward, it becomes increasingly important to continue to collect internationally relevant data about parenting and child outcomes. If recommendations are to be made for policies, interventions, or treatments, such recommendations must be based on accurate balancing of that which is

universal against that which is unique to particular cultural contexts. Thus it is necessary to employ statistical methods that are able to adequately and accurately analyze data across countries.

As I will outline below—and is evident in the literature (Kreft & de Leeuw, 1998; Luke, 2004; Rabe-Hesketh & Skrondal, 2012; Raudenbush & Bryk, 2002; Singer & Willett, 2003)—multilevel models are eminently suited for cross-cultural research in that they are not only able to *control* for the clustering of study participants within countries, but are also able to *explore the variation*—or *consistency*—of patterns of family life across countries.

Long ago, Cesaire, writing about liberatory movements wrote...

“My conception of the universal is that of a universal enriched by all that is particular, a universal enriched by every particular: the deepening and coexistence of all particulars.” (Cesaire, 1956)

It is this sensibility that I hope to echo in my discussion of the multilevel model below.

2.6 Multilevel Models As The Study Of Variation

“Every being cries out silently to be read differently.”

— Simone Weil, *Gravity and Grace* as reported in Su (2017)

Multilevel models are sometimes seen as an analytic technique that *controls* for the clustering or nesting of individuals inside larger social units such as schools, neighborhoods, or countries. I will describe below how this ability to *control* for clustering is indeed an important and crucial aspect of multilevel models.

However, my argument here is that multilevel models are better seen as a method to *explore* the variation in inherent within nested or clustered data. Again, while these issues are well understood within the statistical literature (Kreft & de Leeuw, 1998; Luke, 2004; Rabe-Hesketh & Skrondal, 2012; Raudenbush & Bryk, 2002; Singer & Willett, 2003), they are less often noted in applied research.

In the graph below, imagine that physical punishment, or some other risk factor, is associated with detrimental mental health outcomes. Each country in the data has its own *country specific regression line*.

Plausible Alternative Patterns of Between Country Vari In The Relationship of Physical Punishment With Psychological Wellbeing

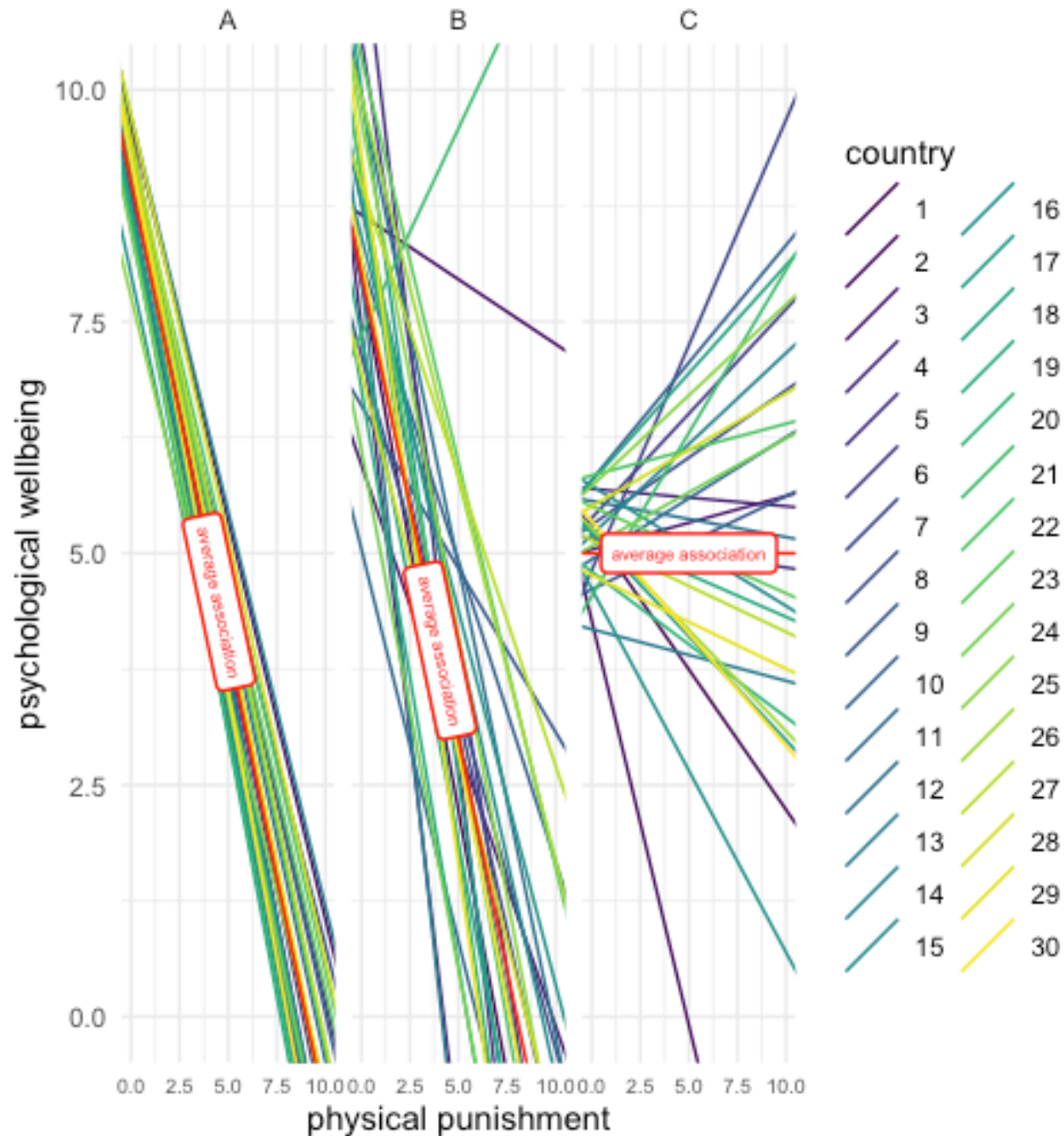


Figure 2.1: Plausible Alternative Patterns of Between Country Variation

In Panel A, there is some variation in the *intercept*, which is equivalent to saying that there is some variation in the average level of psychological well-being across countries. When we look at the slope of the country-specific regression lines in Panel A, we notice that there is little

variation in these *slopes*. Put another way, there is a great amount of consistency in the slopes of the country-specific regression lines: parental use of physical punishment is consistently associated with decreases in child psychological wellbeing across countries.

In Panel B, the situation is different. There is more variation in the *intercept*, that is, more variation between countries in the initial or average amount of psychological well-being. There is also more variation in the *slopes* of the country-specific regression lines. While the average association between physical punishment and psychological well-being is very similar to that in Panel A, there is more variation across countries, in the relationship of physical punishment and child psychological wellbeing, which would likely merit exploration were one considering developing programs, policies or interventions for different countries.

Lastly, the pattern of variation in Panel C is considerably different from either Panel A or Panel B. The average association of physical punishment with psychological well-being in the hypothetical scenario represented by Panel C is approximately 0. There is some variation in the *intercepts* of the country-specific regression lines. Additionally, there is considerable variation in the *slopes* of the country-specific regression line, suggesting that the use of physical punishment might be beneficial in some countries, and detrimental in others.

Empirically, data generally suggest a scenario somewhere between Panel A and Panel B, but these different hypothetical scenarios afford us the opportunity to think about possible patterns of variation.

A second pedagogically helpful example might be obtained if we flip the slopes in the diagram, and consider a different set of independent variables, perhaps some kind of treatment or intervention designed to improve psychological well-being.

Considering an Intervention or Treatment Across Coun

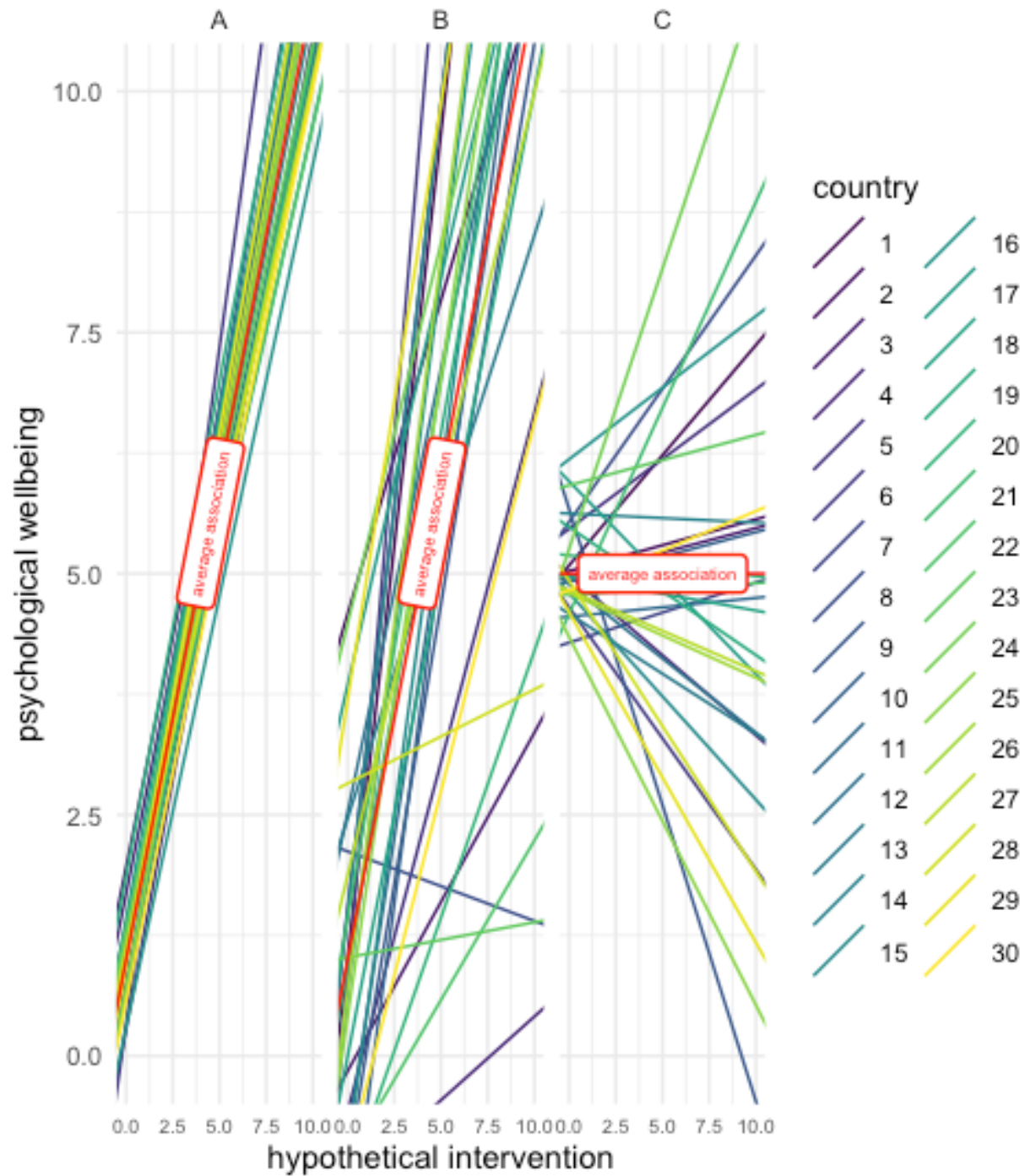


Figure 2.2: Considering an Intervention or Treatment Across Countries

We see a similar pattern as before, but the use of a different substantive example may be illustrative.

In Panel A, there is relative consistency in the initial levels of psychological well-being across countries, as well as consistency in the degree to which the intervention is associated with improvements in psychological well-being across countries.

In Panel B, we see more variation in both initial levels of psychological well-being, but also more variation in the association of the intervention with improvements in psychological well-being.

Lastly, in Panel C, we note an overall association of the intervention with psychological well-being that is close to zero. However associations vary widely by countries. In some countries there appears to be evidence that the intervention is beneficial, while in other countries there appears to be evidence that the intervention is not beneficial, or even possibly harmful.

Thus, I emphasize an approach to multilevel modeling that sees multilevel modeling as the *study of variation*, not simply *accounting for variation*, or *controlling for variation*.

“...universal theorizing requires adequately sampled (i.e., diverse) data and better appreciation of issues of comparability and the most powerful theories ought to predict and explain variation, not sweep variation under the rug.” (Blasi et al., 2022)

Again, sophisticated treatments of all of the ideas are available in one form or another across the excellent textbooks on multilevel modeling (Kreft & de Leeuw, 1998; Luke, 2004; Rabe-Hesketh & Skrondal, 2012; Raudenbush & Bryk, 2002; Singer & Willett, 2003). However, some of these ideas appear less often in applied research, and my intention here is to make the application of these ideas to applied research more clear.

3. Simulated Multi-Country (Multilevel) Data

I use simulated data in this example. Data come from 30 hypothetical countries. Data contain measures of a few key aspects of parenting that have proven salient in the empirical literature on parenting to date: parental warmth, and physical punishment. Both parenting measures are normally distributed variables. Our outcome is conceptualized as a positive mental health outcome or behavioral outcome, and higher levels of outcome are considered to be better. Statistically, the data are clustered within countries.

In this simulation, I construct the data so that warmth is positively related to the outcome, while physical punishment is negatively related to the outcome.

Table 3.1: Simulated Multilevel Data

id	country	warmth	physical_punishment	outcome
1	15	129.6	84.86	685.3
2	24	106.2	90.91	330.5
3	11	83.41	95.94	-130.8
4	8	90.59	96	100.5
5	21	95.06	85.3	149.1
6	20	66.6	102.3	238.1

Desirable Mental Health Outcome by Parental Warmth By Country

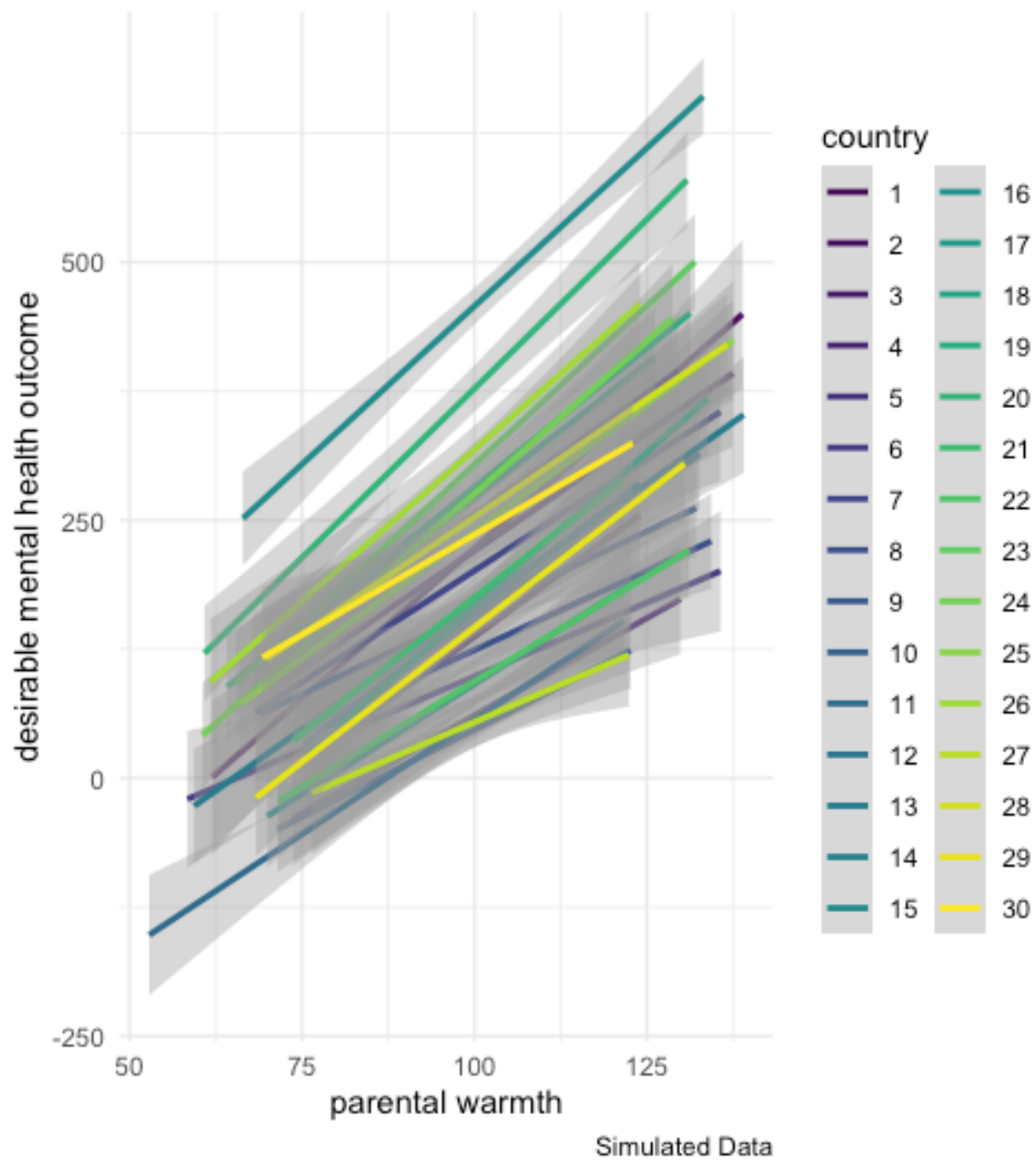


Figure 3.1: Graph of Simulated Data

4. Software

In this document, I use Stata (StataCorp, 2021b) to analyze data. Stata is my software of choice in this document because of Stata's overall ease of use and intuitiveness. The creators of Stata have created a powerful program that is extremely simple to use, but with a wide range of both basic and advanced statistical capabilities.

The general idea of most Stata commands is:

```
do_something to_a_variable_or_variables, options
```

Often it is not necessary to use any options since the authors of Stata have done such a good job of thinking about the defaults.

For the sake of illustration, a few Stata commands are listed below.

Table 4.1: Example Stata Commands

Task	Command
Open data	<code>use mydata.dta</code>
Descriptive statistics	<code>summarize x y</code>
Frequencies	<code>tabulate x</code>
Correlation	<code>corr x y</code>
Regression	<code>regress y x</code>
Logistic Regression	<code>logit y x, or¹</code>
Multilevel Model	<code>mixed y x group:</code>

It is this multilevel syntax, `mixed y x || group:` that we will be using throughout this document.

5. Conceptual Framework

5.1 Units of Analysis

When confronted with multilevel data, one has a number of choices about the units of analysis: one could consider individuals to be the units of analysis; or, one could consider the larger social units to be the units of analyses. With multilevel analytic methods, one is able to avoid this false dichotomy, and to conceptualize the data from a multilevel perspective, wherein both individuals and social units are different levels of the same analysis. I discuss some of the statistical considerations of different ideas about the units of analysis in [Section 6.8](#)

Figure 5.1: Conceptual Framework

¹ Here we use the `,or` option to ask for *odds ratios* instead of *logit coefficients*.

5.2 Variables and Processes at Multiple Levels

Table 5.1: Multiple Levels of Variables

conceptuallevel	statisticallevel1	statisticallevel2
1	Individual response about parenting or mental health	Aggregated responses about parenting or mental health
2	Individual response about community	Aggregated response about community
2	N/A	Administrative indicator of social unit

In this document, I distinguish between *conceptual* and *statistical* levels.

By *conceptual* level, I refer to whether a variable is *conceptualized* to be measure of an *individual* level characteristic, such as parenting or mental health, or a *community* level construct, such as community collective efficacy, or community safety.

By *statistical* level, I refer to whether a variable measures an *individual* response, or an *aggregated* response.

- Thus, $\text{mental health}_{ij}$ or parenting_{ij} would be considered in the terminology that I am using to be a variable both *conceptually* and *statistically* at Level 1.
- $\overline{\text{mental health}}_j$ or $\overline{\text{parenting}}_j$ would be variables that *conceptually* come from Level 1 responses, but are *statistically* aggregated to Level 2.
- Using my terminology, $\text{community collective efficacy}_{ij}$ or $\text{community safety}_{ij}$ would be considered to be a variable both *conceptually* at Level 2, but *statistically* at Level 1.
- $\overline{\text{community collective efficacy}}_j$ or $\overline{\text{community safety}}_j$ would be variables that *conceptually* come from Level 2 responses, but are *statistically* aggregated to Level 2.

Some variables only exist at Level 2, and their Level 1 counterparts are undefined. For example, the size of a school, neighborhood, or country, is inherently a Level 2 variable, with no Level 1 counterpart. Similarly, some administrative indicators, such as the Gini level of inequality, while developed by calculating across Level 1 responses, have no easily definable Level 1 counterpart.

6. The Cross Sectional Multilevel Model

6.1 The Equation

The equation for the multilevel model can be written in several ways: as multiple levels of equations; or as a single equation. The advantage of having multiple levels of equations is that these multiple equations make clear the multiple levels of the data, and thus conform to an initial understanding of how a multilevel model should be estimated. However, results from multiple levels of equations quickly become difficult to interpret, and thus, I will not spend a great deal of time on discussing empirical results of the two level formulation. Whether multiple levels of equations, or a single equation are employed, the numerical results are the same.

6.1.1 Two Levels of Equations

I start with two levels of equations: Level 1 at the level of the individual; and Level 2 at the level of the country.

6.1.1.1 Level 1 (Individuals)

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + \beta_{2j}z_{ij} + \beta_{3j}t_{ij} + e_{ij} \quad (6.1)$$

6.1.1.2 Level 2 (Countries)

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (6.2)$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

$$\beta_{2j} = \gamma_{20}$$

$$\beta_{3j} = \gamma_{30}$$

Here y_{ij} is the dependent variable, or outcome for the model. We note that the ij subscripts indicate that this is outcome y for individual i in country j . Note that the outcome is at Level 1, or the level of individuals. β_{0j} is a regression intercept, and the other β 's² are regression slope parameters. x_{ij} and z_{ij} are independent variables and t_{ij} is an independent variable indicating the time at which different data points are measured. I note that in this discussion I am *not* considering a model in which there are repeated observations on the same individuals, although the multilevel model is certainly extensible to such cases. u_{0j} is a random intercept for the β_{0j} term, and u_{1j} is a random slope for the β_{1j} term, indicating that we are modeling cross country variation in these parameters. The other β terms are not modeled as having random country level variation, although this could certainly be a possibility in subsequent models.

In this formulation of the multilevel model, each regression parameter β in the level 1 equation is the outcome of an equation at Level 2.

6.1.2 One Level of Equations

By simply substituting the values of the Level 2 equations into the Level 1 equations, we obtain:

$$y_{ij} = \beta_0 + \beta_1x_{ij} + \beta_2z_{ij} + \beta_3t_{ij} + u_{0j} + u_{1j} \times x + e_{ij} \quad (6.3)$$

Here again y_{ij} is the dependent variable, or outcome for the model. β_0 is a regression intercept, and the β 's are regression parameters. x_{ij} and z_{ij} are independent variables and t is an independent variable indicating the time at which different data points.

² Technically, all of these β 's could be written as β_j since the multilevel model could be said to estimate a regression parameter for each group, in this case each country. One could even write β_{jk} to represent the regression parameter for the k^{th} independent variable the for the j^{th} group or country. To keep matters simple, I simply write β in most cases.

While I purposely strive to keep the discussion of multilevel modeling ideas somewhat general in this document, I generally think of y_{ij} as a desirable or “good” outcome. x is often conceptualized as a protective factor, while z is often conceptualized as a risk factor.

Thus, phrased less abstractly, and more concretely in terms of the ideas of this document:

$$\text{outcome}_{ij} = \beta_0 + \beta_1 \text{parental warmth}_{ij} + \beta_2 \text{physical punishment}_{ij} + \beta_3 \text{time}_{ij} + u_{0j} + u_{1j} \times \text{parental warmth} + e_{ij} \quad (6.4)$$

Drawing upon ideas from [Chapter 4](#), this single level equation can be easily represented in Stata syntax.

```
mixed outcome warmth physical_punishment || country: warmth
```

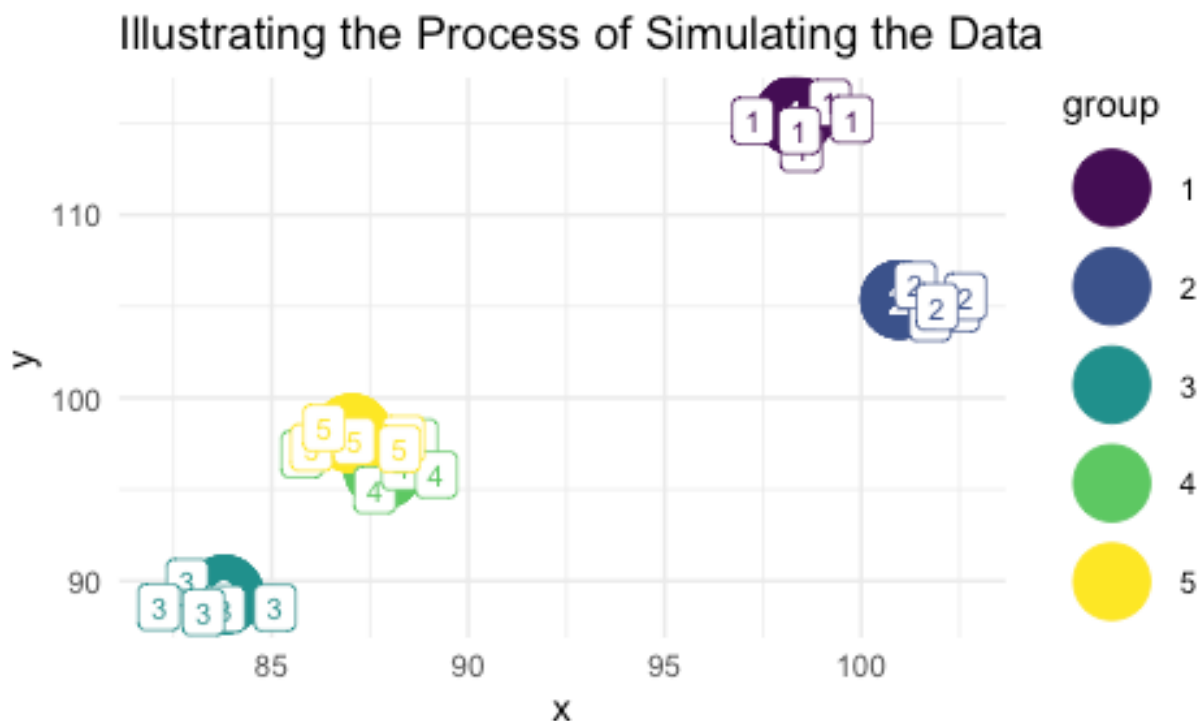
6.2 Estimating Standard Errors And p Values

6.2.1 Introduction

If the data are grouped, nested, or clustered, then this aspect of the structure of the data needs to be accounted for. Bland & Altman (1994) describe a simulation in which grouped data are artificially generated according to the following procedure.

“The data were generated from random numbers, and there is no relation between X and Y at all. Firstly, values of X and Y were generated for each ‘subject,’ then a further random number was added to make the individual observation.” (Bland & Altman, 1994)

The graph below illustrates the process of simulating the data.



6.2.2 Compare OLS and MLM

An analysis that is not aware of the grouped nature of the data will give biased results, will mis-estimate standard errors, and importantly, will often attribute statistical significance to some of the independent variables when this is not appropriate (Bland & Altman, 1994; Raudenbush & Bryk, 2002).

In the example below, we compare a simple ordinary least squares analysis of the data with a multilevel model that accounts for the clustered nature of the data.

The Stata syntax that we use for each analysis is:

- `regress y x`
- `mixed y x || group:`

	OLS		MLM
x	1.046	**	0.039
Intercept	4.488		97.005 **
var(_cons)			74.523
var(e)			0.594
Number of observations	25		

** p<.01, * p<.05

We see that in the ordinary least squares analysis, the independent variable is judged to have a statistically significant association with the dependent variable. The more appropriate multilevel model finds that in fact the independent variable x is *not* associated with y . Thus, the multilevel model provides more accurate results than OLS in the presence of clustered data.

6.3 Multilevel Structure

Associations between two variables can be *very different* (or even *reversed*) depending upon whether or not the analysis is “aware” of the grouped, nested, or clustered nature of the data (Gelman et al., 2007). In the example presented here, the groups are countries, but could as easily be neighborhoods, communities, or schools.

For pedagogical purposes, I use an example with very few clusters, although it would be more appropriate to apply multilevel analysis to an example with many more clusters e.g. ($N_{\text{clusters}} \geq 30$)

A model that is “aware” of the clustered nature of the data may provide very different—likely better—substantive conclusions than a model that is not aware of the clustered nature of the data.

6.3.1 Use Some Data Simulated For This Particular Example

6.3.2 Graphs

6.3.2.1 A “Naive” Graph

This “naive” graph is unaware of the grouped nature of the data. Notice that the overall regression line slopes downward, even though there is some suggestion that *within each group* the regression lines may slope upward.

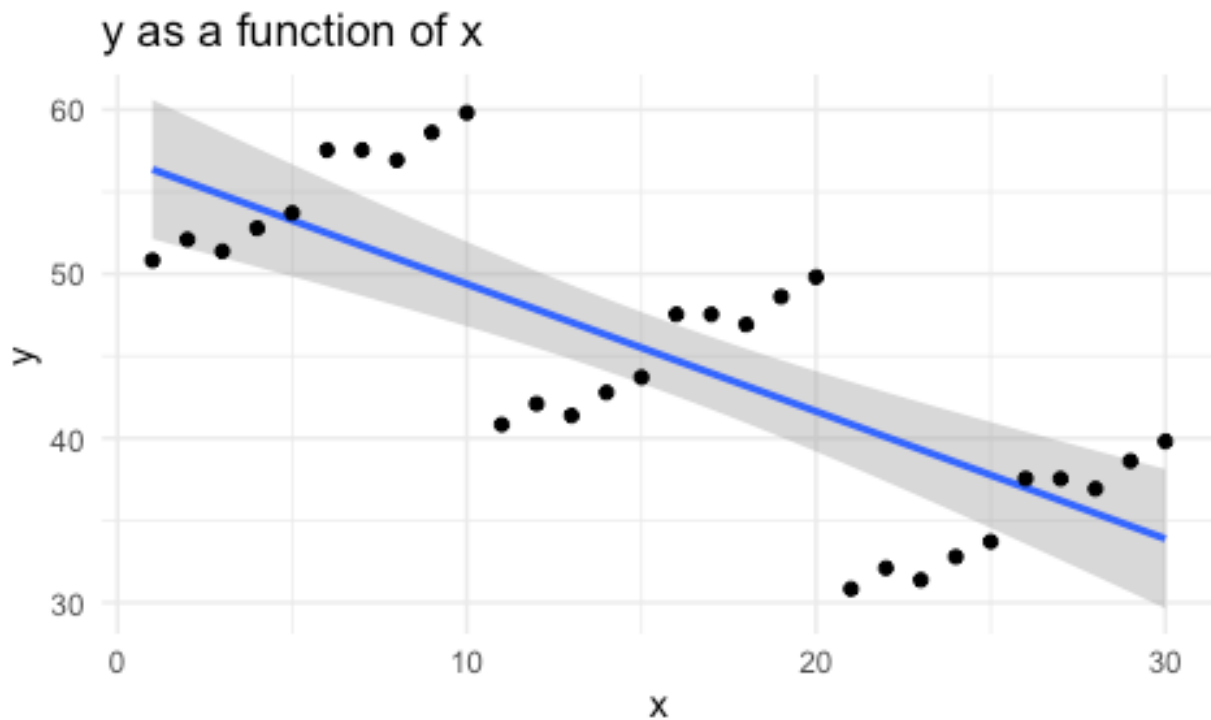


Figure 6.1: A ‘Naive’ Graph

6.3.2.2 An “Aware” Graph

This “aware” graph is aware of the grouped nature of the data. The graph is “aware” of the grouped or clustered nature of the data, and provides indication that the regression lines *when accounting for group* slope upward.

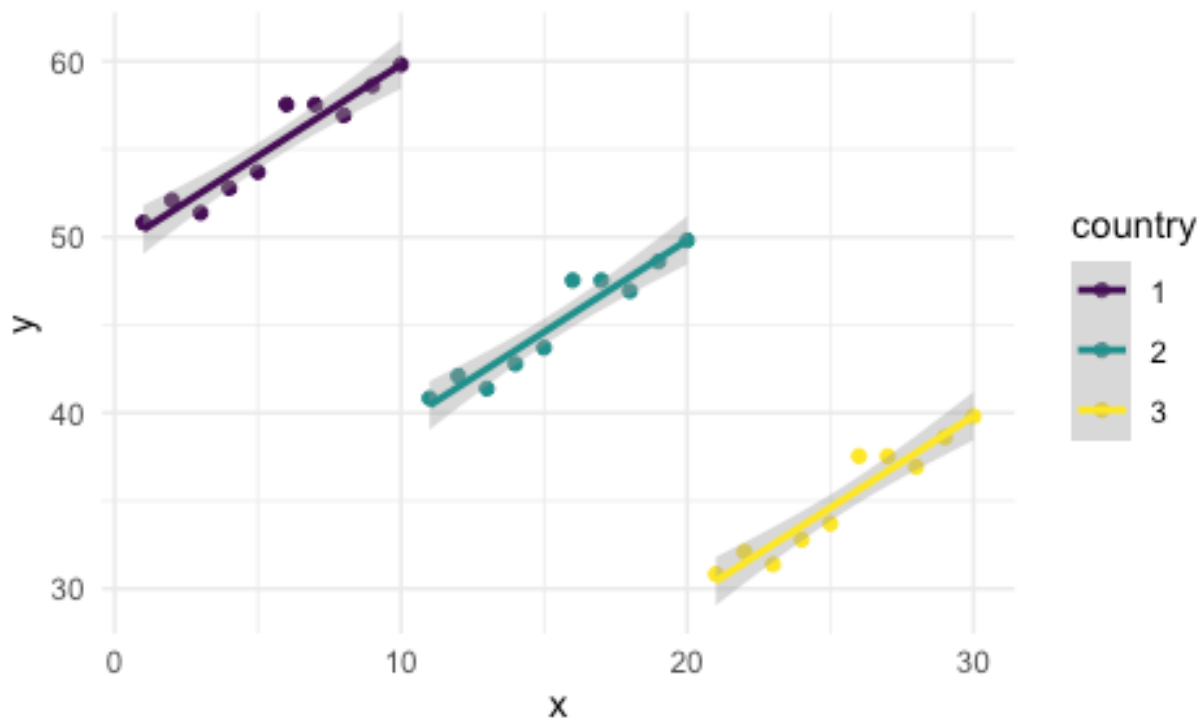


Figure 6.2: An 'Aware' Graph

6.3.3 Regressions

6.3.3.1 A "Naive" OLS Analysis vs. An "Aware" MLM Analysis

The Stata syntax that we use for these analyses is:

- `regress y x`
- `mixed y x || country:`

The OLS model with only x as a covariate is not aware of the grouped structure of the data, and the coefficient for x in the OLS model reflects this. The coefficient for x in the OLS model is *negative*, and statistically significant.

The multilevel model is aware of the grouped structure of the data, and the coefficient for x in the multilevel model is *positive*, and statistically significant.

	OLS		MLM	
x	-0.771	**	1.072	**
Intercept	57.122	**	28.553	**
var(_cons)			286.088	
var(e)			0.758	
Number of observations	30			

** $p < .01$, * $p < .05$

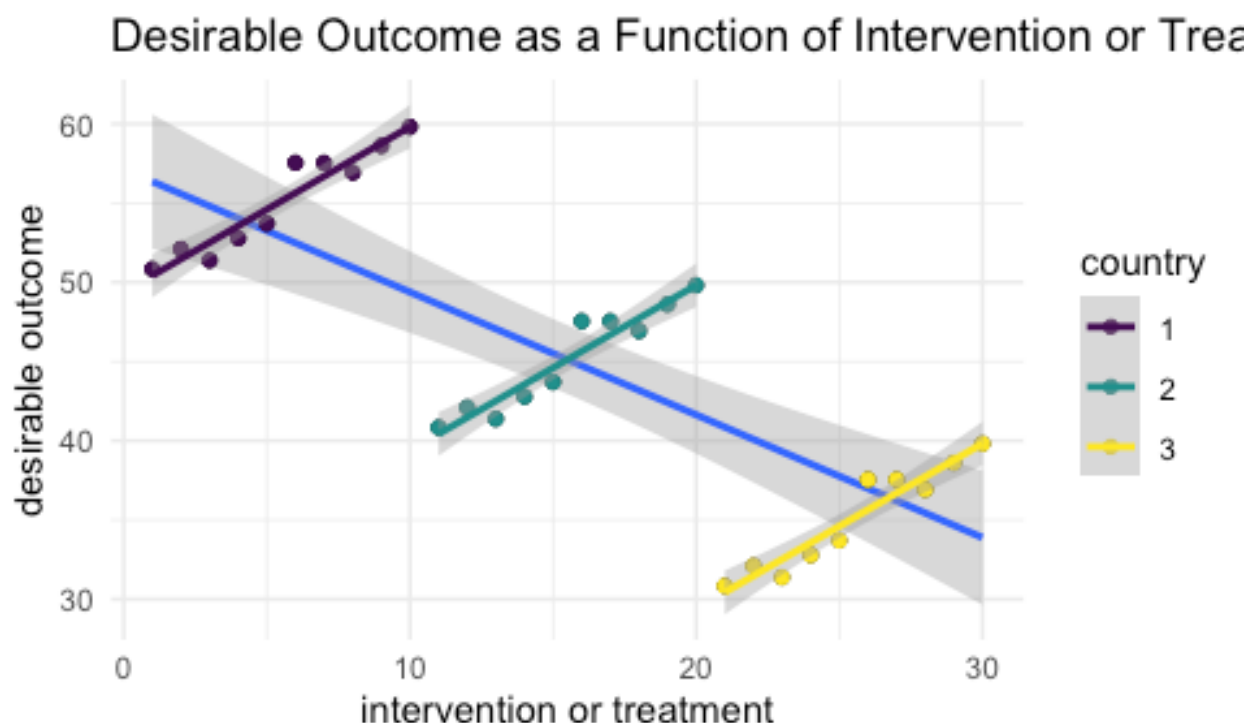
6.3.4 A Thought Experiment

When might a situation like this arise in practice? This is surprisingly difficult to think through.

Imagine that x is a protective factor, or an intervention or treatment. Imagine that y is a desirable outcome, like improved mental health or psychological well being.

Now imagine that residents of countries provide more of the protective factor or more of the intervention in situations where there are lower levels of the desirable outcome. If one thinks about it, this is a very plausible situation.

A naive analysis that was unaware of the grouped nature of the data would therefore misconstrue the results, suggesting that the intervention was harmful, when it was in fact helpful.



A Heuristic Example

These data are constructed to provide this kind of extreme example, but it is easy to see how multilevel thinking, and multilevel analysis may provide better answers than one would get if one ignored the grouped nature of the data.

6.4 Regression With Simulated Multi-Country Data

Let's examine the results of a multilevel regression with the simulated multicountry data. I will again imagine that the desirable outcome is an outcome such as improved psychological wellbeing.

The Stata syntax that we use is:

```
mixed outcome warmth physical_punishment || country: warmth
```

	cross_sectional	
warmth	4.855	**
physical_punishment	-3.242	**
Intercept	30.930	
var(warmth)	0.800	
var(_cons)	1018.485	
var(e)	2574.967	

** p<.01, * p<.05

The data suggest that parental warmth is positively associated with the desirable outcome, and that this result is statistically significant. Parental use of physical punishment is associated with statistically significant decreases in the desirable outcome. I note that there is some variation in the *constant* indicating that there is some variation in the initial or average levels of the desirable outcome—again improved psychological well-being—that is attributable to country.

There is also some variation in the slope associated with parental warmth that is attributable to country. Thus, while largely consistent, the relationship of parental warmth with child outcomes differs somewhat from country to country.

6.5 Summary of Advantages Of The Multilevel Model

The discussion so far gives an idea of the advantages of the multilevel model for studying intrinsically multilevel data: children in classrooms or schools; individuals or families in neighborhoods; individuals or families in countries. These advantages can be summarized below:

1. Standard errors are estimated correctly as is statistical significance. This means that p values are correctly estimated accounting for the clustered or nested nature of the data. More colloquially, this most often means that we do not make the mistake of attributing statistical significance to a given risk or protective factor, when such a statistical significance is not warranted. Put even more straightforwardly correct estimation of standard errors and statistical significance prevents us from seeing results that are simply not present in the data, whether those concern risk factors or protective factors.
2. Regression coefficients are estimated correctly accounting for the clustered or nested structure of the data. If one does not account for the clustered or nested structure of the data, regression slopes can be estimated as negative when they are more correctly estimated as positive, or as null, or conversely estimated as positive when there are more correctly seen as negative (or null). Again, to phrase things in a more colloquial fashion, this means that we do not judge something to be a risk factor when it is in fact a protective factor or a null effect; or a protective factor when it is in fact a risk factor, or a null effect.
3. An increasing focus of statistical estimation is not to focus on particular regression parameters, but instead to predict outcomes for particular combinations of independent variables. Predictions from a multilevel model could be said to be best predictions in that groups are weighted by their precision, contributing to an estimate which makes better predictions than would a simple average. More colloquially, multilevel models allow us to predict outcomes better and more accurately than would be possible with simple or more naïve models.

6.6 Predicted Values

According to “**Stein’s Paradox**”, predictions from a multilevel model may be better than the mean.

shrinkage

6.7 Variation

Above, in [Section 2.6](#), I have referred to multilevel models as the study of variation. Now that I have provided some discussion of the multilevel model, more statistical “unpacking” of ideas about variation is warranted.

I provide again, for pedagogical purposes, the example substantive equation that I have been using in this document.

$$\text{outcome}_{ij} = \beta_0 + \beta_1 \text{parental warmth}_{ij} + \beta_2 \text{physical punishment}_{ij} + \beta_3 \text{time}_{ij} + u_{0j} + u_{1j} \times \text{parental warmth} + e_{ij} \quad (6.5)$$

6.7.1 Measured and Unmeasured Variation

The first thing to note about the equation is that it can be divided into measured and unmeasured variation.

This is most easily seen if we introduce the idea of an unconditional model.

6.7.2 Unconditional Model

The unconditional model is a model with no x ’s or covariates.

$$\text{outcome}_{ij} = \beta_0 + u_{0j} + e_{ij} \quad (6.6)$$

Here, outcome_{ij} is a function of an intercept β_0 , a country specific error term, u_{0j} , and an individual level error term e_{ij} .

Thus, all of the variation in outcome_{ij} is—given the unconditional nature of our model—due to unmeasured variation at the country and individual level.

6.7.3 Intra-Class Correlation Coefficient

We then introduce a measure known as the Intra-Class Correlation Coefficient, (ICC) that can be computed from this unconditional model.

$$\text{ICC} = \frac{\text{var}(u_{0j})}{\text{var}(u_{0j}) + \text{var}(e_{ij})} \quad (6.7)$$

Heuristically:

$$\text{ICC} = \frac{\text{group level variation}}{\text{group level variation} + \text{individual level variation}} \quad (6.8)$$

The ICC from the *unconditional* model ([Equation 6.6](#)) is the most informative ICC as it represents the amount of variation in the dependent variable that could *potentially* be explained by the grouping variable.

As we add covariates, x 's, to the model the ICC will most often decrease.

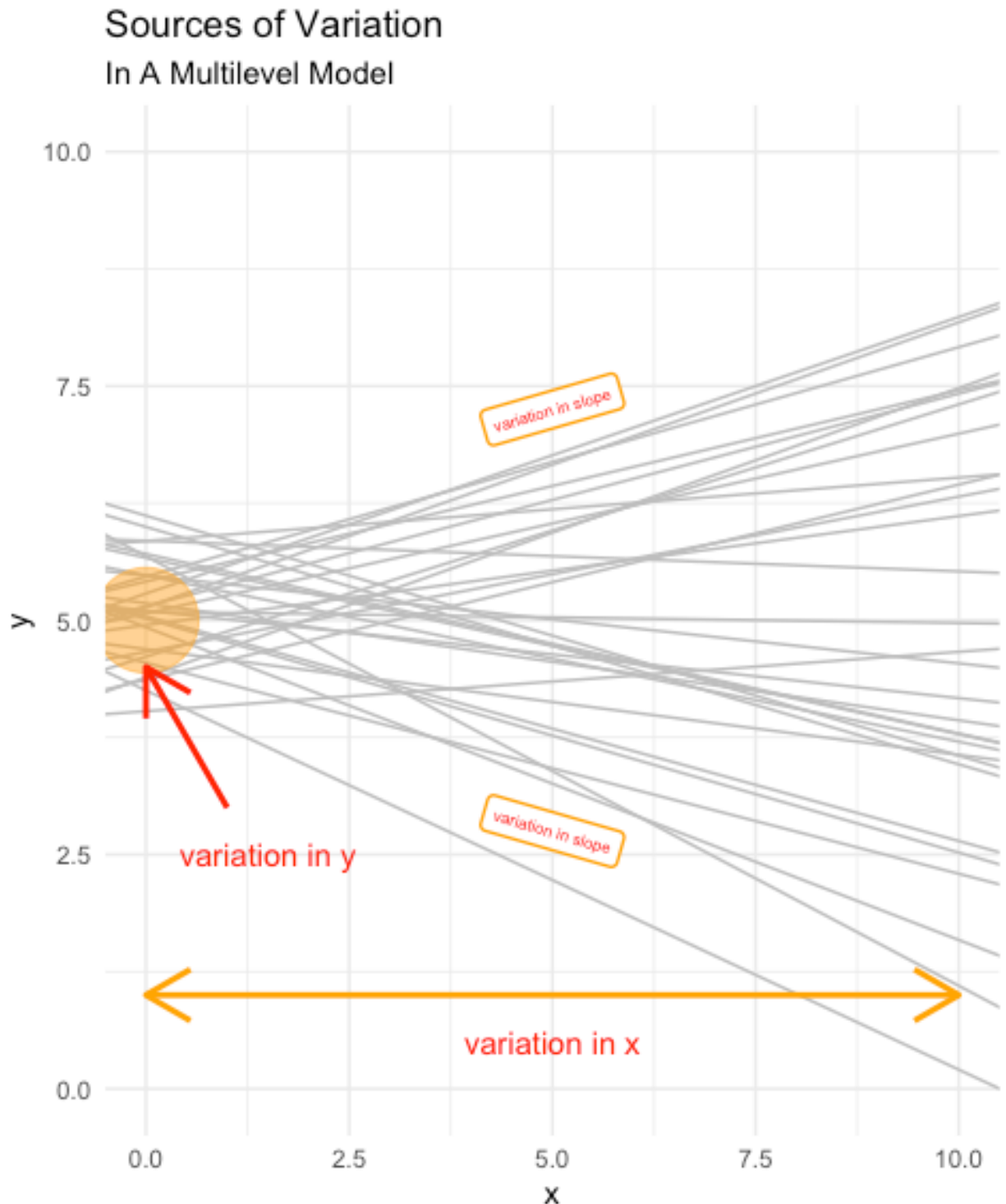


Figure 6.3: Sources of Variation in a Multilevel Model

6.7.4 Variation In Intercepts or Outcomes

In [Equation 6.5](#), $\text{var}(u_{0j})$ is the model estimated amount of variation in the *outcome*, y_{ij} .

In the regression in [Section 6.4](#), there is discernible between country variation, but more of the variation is between individuals within the same country. Put another way, there is a moderate tendency for children in families in the same country to have similar outcomes, but two children in families in the same country may also have very different outcomes. Children from families in different countries may be as similar as children from families in the same country.

6.7.5 Variation In Predictors

Equally important, I think, but much less frequently explored than variation in *outcomes*, is the possibility of variation in *predictors*. In the substantive example that we have employed so far, the *predictors* are different *parenting behaviors*, so considering variation in *predictors* allows us to consider variation in *parenting behaviors*, as well as variation in the *outcomes* of those behaviors.

We would estimate variation in behaviors in much the same way that we would estimate variation in outcomes, estimating an unconditional model, but substituting x for y .

$$x_{ij} = \beta_0 + u_{0j} + e_{ij} \quad (6.9)$$

Then, similarly, the variation in a predictor attributable to the clustered nature of the data—in this case the clustering of individuals in countries—is given by:

$$\text{ICC}_x = \frac{\text{var}(u_{0j})}{\text{var}(u_{0j}) + \text{var}(e_{ij})} \quad (6.10)$$

6.7.6 Variation in Slopes

Another possible type of variation to investigate is variation in the relationship of x and y , which is represented in the multilevel model by examining variation in the β 's.

6.7.7 Variation As An Outcome

Even less common is to examine *variation* itself as an outcome (Burkner, 2018).

$$\sigma_{yij} = \beta_0 + \beta_1 x_1 + u_{0j} + e_{ij} \quad (6.11)$$

Here, the variation in the outcome, σ_{yij} , rather than the mean level of the outcome, y_{ij} , is the focus of interest. My notation for [Equation 6.11](#) draws upon Burkner (2018)'s notation, but is modified in order to be consistent with the rest of this document.

Why might such models be of conceptual interest? Imagine for example, that the *variation* in psychological well-being is higher in countries with higher levels of poverty, or higher levels of income inequality. The use of such models as this, discussed in more detail by Burkner (2018), would allow us to explore such a question.

Of note, while I do not explore in detail differences between Bayesian and frequentist approaches to multilevel modeling in this document, these models are likely to be only estimable with Bayesian software rather than with frequentist software (Burkner, 2018).

6.7.8 Maximal Models

Hypothetically, one might imagine that there could be group level unobserved factors which affect regression slopes—i.e. the relationship between a predictor x and outcome variable y —arguably, were one to ignore these unobserved factors in statistical estimation, they would show up either in an error term, or in the regression coefficients themselves. Were they to show up in the regression coefficients this would represent statistical bias and a substantive mis-estimation of important effects. thus, there is a conceptual argument for including as many random effects—i.e. random slopes—in a statistical model as possible.

Models with all possible random effects are termed *maximal models* (Barr et al., 2013; Frank, 2018). Such models include a large number of random slopes, e.g. u_1, u_2, u_3, \dots , etc. even when some of those estimated slopes are close to 0. Such models may be more easily estimable when using Bayesian estimation (Frank, 2018), a topic which I do not cover in detail in this document.

It should be noted that Matuschek et al. (2017) argue that such a *maximal* approach may lead to a loss of statistical power and further argue that one should adhere to “a random effect structure that is supported by the data.” In contrast, Nalborczyk et al. (2019) argue that maximal models are supported under the Bayesian approach. Oberauer (2022) also argues for including multiple random slopes. Schielzeth & Forstmeier (2009) make a similar argument from a frequentist perspective.

6.8 Some Wrong (or Partially Wrong) Approaches

When data are clustered—e.g. residents in neighborhoods, children in schools, families in countries—it is worth discussing the fact that we have several choices statistically as how to proceed, other than using a multilevel model. Given the discussion so far, we can see the advantages of a multilevel model over these other approaches:

1. First, we could simply ignore the clustering, and treat the data as though it were composed of statistically independent individuals, i.e. statistically independent e_i . As we have discussed above, however, this approach has at least two disadvantages. First, as discussed in [Section 6.2](#), this approach will mis-estimate standard errors, most often underestimating them, resulting in underestimated p values and false positives. Second, as discussed in [Section 6.3](#) ignoring clustering runs the risk of estimating regression β 's that are not estimated with information about the multilevel structure of the data, with the possibility that β coefficients may not only have incorrect statistical significance, but also incorrect magnitude, and even incorrect sign.
2. A second approach would be to *aggregate* the data to the level of the higher social unit, e.g. aggregating the data at the level of the neighborhood. Here we run into an idea similar to that discussed in [Section 6.3](#), the “ecological fallacy”: the idea that group level and individual level relationships are necessarily the same (Firebaugh, 2001).
3. Lastly, we could adopt a statistical strategy of *clustering* the standard errors. Clustering the standard errors means that standard errors are corrected for the non-independence of the e_i within clusters. Thus, p values are estimated correctly. However, clustering still does not account for the multilevel structure of the data ([Section 6.3](#)), and thus when relationships between x 's and y at different levels of the data are very different, simply clustering the standard errors may not give correct estimates of the β 's.

7. The Longitudinal Multilevel Model

“Mathematics is the art of giving the same name to different things.” (Poincare, 1908)

Counter-intuitively, and surprisingly, the mathematics of estimating models with cross-sectional clustered data easily generalizes to longitudinal data. In cross sectional clustered data, we imagine *individuals clustered in neighborhoods, schools, or countries*.

Table 7.1: Levels in Cross-Sectional Data

Level	Example(s)
1	Individuals
2	Schools Neighborhoods Countries

In longitudinal data, we consider the *first level* to be that of *time points*, or *study waves*, which we sometimes call the *person-observation*. The *second level* is then the individual.

Table 7.2: Levels in Longitudinal Data

Level	Example(s)
1	Timepoints
2	Individuals

While it is less common, we could then easily add additional clustering to this longitudinal model, for example, clustering of individuals inside social units.

Table 7.3: Multiple Levels in Longitudinal Data

Level	Example(s)
1	Timepoints
2	Individuals
3	Schools Neighborhoods Countries

7.1 Use Data With Multiple Observations Per Individual

Multilevel data suitable for longitudinal analysis therefore has *multiple rows of data per individual*. Put another way, *every row of data is a person-timepoint*.

This method of organizing data is known as the *long* format. Another way of organizing longitudinal data—which I do not discuss in detail here—is the *wide* format in which every individual still has only one row of data. In *wide* data, the different timepoints are in different columns of data.

Table 7.4: Data in Long Format

id	t	x
1	1	10

id	t	x
1	2	20
1	3	30
2	1	20
2	2	30
2	3	40

Table 7.5: Data in Wide Format

id	x1	x2	x3
1	10	20	30
2	20	30	40

Table 7.6: Simulated Longitudinal Multilevel Data

id	warmth	physical_punishment	country	outcome	t	x_noise
1	129.6	84.86	15	685.3	1	2.872
1	124.1	84.86	15	675.8	2	-12.67
1	126.9	84.86	15	691.2	3	-19.88
2	106.2	90.91	24	330.5	1	-6.262
2	128.3	90.91	24	479.6	2	5.793
2	142.2	90.91	24	580.6	3	9.695

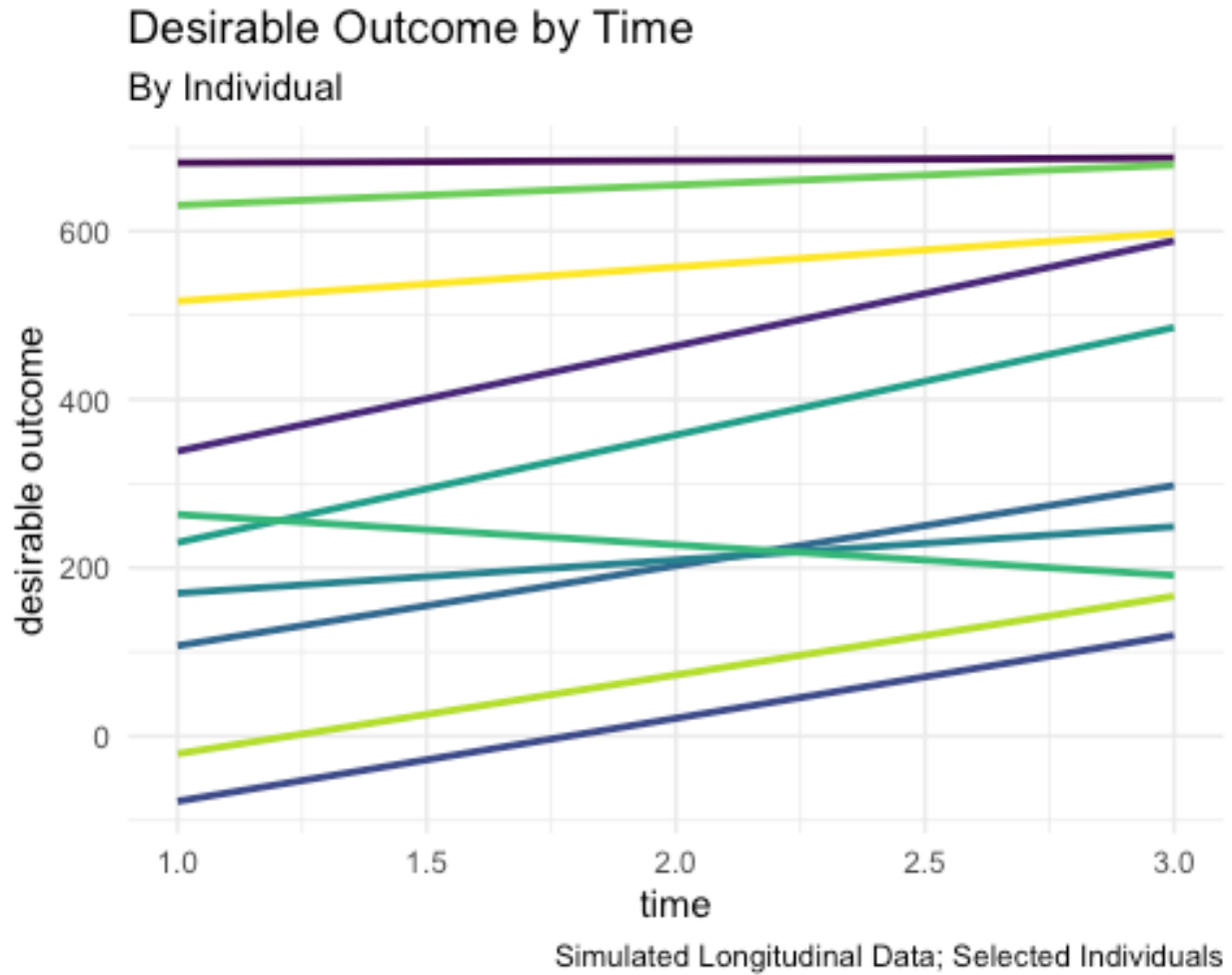


Figure 7.1: Graph of Simulated Longitudinal Data

7.2 The Equation

When data are in *long* format, the following equation is applicable.

$$\text{outcome}_{itj} = \beta_0 + \beta_1 \text{parental warmth}_{itj} + \beta_2 \text{physical punishment}_{itj} + \beta_3 \text{time}_{itj} + \quad (7.1)$$

$$u_{0j} + u_{1j} \times \text{parental warmth}_i +$$

$$v_{0i} + v_{1i} \times t + e_{ij}$$

Here I include a random slope (u_{1j}) at the country level for parental warmth, as well as a random slope (v_{1j}) at the individual level for time.

7.3 Regression With Simulated Multi-Country Longitudinal Data

The Stata command that we use to analyze this data is:

```
mixed outcome t warmth physical_punishment || country: warmth || id: t
```

	longitudinal	
t	4.124	**
warmth	4.874	**
physical_punishment	-3.054	**
Intercept	5.860	
var(warmth)	0.890	
var(_cons)	415.073	
var(_cons)	0.000	
var(t)	0.000	
var(e)	2660.874	

** p<.01, * p<.05

Examining the regression results, the results of the model suggest that child outcomes improve over time. Better child outcomes are again associated with parental warmth, and parental use of physical punishment is associated with reduced child outcomes.

7.4 Autocorrelation

When data are ordered by a time variable t , it is possible that observations that are closer together in time will have a higher correlation than observations that are distant in time. In the simplest example, $e_{i,t=k}$ may be correlated with $e_{i,t=k-1}$. This phenomenon is known as *autocorrelation*. As Hooper (2022) would suggest, it may make sense to assume that the correlation between observations “decays with increasing separation in time”.

Most software programs for multilevel modeling allow one to incorporate measures of autocorrelation so that, e.g., $e_{i,t=3}$ is allowed to be correlated with $e_{i,t=2}$, which in turn can be correlated with $e_{i,t=1}$. More complex autocorrelation structures are usually also possible (StataCorp, 2021a).

7.5 Causality

7.5.1 The Importance of Causal Reasoning

Causal reasoning is sometimes considered to be a statistical—or even overly technical—concern. Arguably, however, whenever one is using research to make recommendations about interventions, or treatments, or policies, one is engaging in some form of causal reasoning (Duncan & Gibson-Davis, 2006).

If one is saying that implementing x would result in beneficial changes in y , one is arguing—at least implicitly—that x is one of the causes of y .

It then behooves one to be explicit about this chain of causal reasoning. For example, to continue one of the substantive examples of this document, if one is going to argue for programs, interventions, or treatments that promote *parental warmth*, or that discourage parental use of *physical punishment* with the aim of improving children’s *mental health*, one must be at least reasonably sure that *parental warmth* and *physical punishment* are *causes* of children’s mental health.

Randomized studies provide the best evidence about the internal validity of causal relationships. However, randomized studies have certain limitations (Diener et al., 2022). First of all—especially in a study with a smaller sample—randomization may not always be perfect, and the control and treatment groups may not be statistically equivalent. Secondly, because of ethical concerns some studies can not be conducted with randomization (Diener et al., 2022). For example, in the study of parenting and child development, children cannot ethically be assigned to parents with different styles of parenting and followed over the long term. Thus, methods that provide rigorous causal estimation with observational methods are necessary (Diener et al., 2022).

Because of the assumed superiority of studies that employ randomization, it is sometimes maintained that *correlation is not causation* and that studies that do not make use of randomization are *only observational* and *correlational*, and that results from observational studies cannot be used to support causal conclusions. However, in an important review (Waddington et al., 2022) suggested that studies using appropriately quantitative methods can provide causally robust conclusions.

In a statement salient for social research, Duncan & Gibson-Davis (2006) point out the logical inconsistency of writing that does not rigorously address causal processes, but then goes on to suggest interventions or treatment or policies:

“Developmental studies are usually careful to point out when their data do not come from a randomized experiment. As with much of the nonexperimental literature in developmental psychology, most of the articles then go on to assert that, as a consequence, it is impossible to draw causal inferences from the analysis. Indeed, much of their language describing results is couched in terms of ‘associations’ between child care quality and child outcomes. It is not uncommon, however, to see these papers make explicit statements about effects, and others draw explicit policy conclusions. For instance, NICHD (1997, 876) stated, ‘The interaction analyses provided evidence that high-quality child care served a compensatory function for children whose maternal care was lacking.’ On the policy side, NICHD (2002c, 199) asserted, ‘These findings provide empirical support for policies that improve state regulations for caregiver training and child-staff ratios.’”

“One cannot have it both ways. Studies that do not aspire to causal analysis should make no claim whatsoever about effects and draw no policy conclusions. At the same time, it would be a terrible waste of resources to conduct expensive longitudinal studies without attempting to use them for causal modeling.”

Lastly, because of their often small samples, randomized studies may have high internal validity, but much lower external validity, or generalization to larger populations (Diener et al., 2022). This issue of generalizability becomes increasingly salient, when we are reminded of the fact that so little social and psychological research has been conducted outside of North American contexts (Draper et al., 2022; Henrich et al., 2010). It is necessary to make use of broadly representative observational data sets, and appropriately sophisticated quantitative methods to make causally robust conclusions from observational data.

7.5.2 Formal Criteria of Causality

For x to be a cause of y , one needs the following 3 things to be true (Holland, 1986).

1. x is (are) associated with (correlated with) y .

2. x come(s) before y in time.
3. z —or other factors—cannot explain the association of (correlation of) x and y .

Figure 7.2: Formal Criteria of Causality

If z is omitted from the regression model, then the estimates for $x \rightarrow y$ (i.e. $\beta_{x \rightarrow y}$) will be biased. In the most common scenario $\beta_{x \rightarrow y}$ will likely be an over-estimate of the effect, and statistical significance of $\beta_{x \rightarrow y}$ may represent a false positive.

It is likely useful to restate the above abstract statements in terms of the substantive issues that I have been considering so far in this document.

For *parenting* to be a cause of *child outcomes*, one needs the following 3 things to be true (Holland, 1986).

1. *parenting* is (are) associated with (correlated with) *child outcomes*.
2. *parenting* come(s) before *child outcomes* in time.
3. *SES, community characteristics*—or other factors—cannot explain the association of (correlation of) *parenting* and *child outcomes*.

Figure 7.3: Formal Criteria of Causality

If *other factors* are omitted from the regression model, then the estimates for *parenting* \rightarrow *child outcome* (i.e. $\beta_{\text{parenting} \rightarrow \text{child outcome}}$) will be biased. In the most common scenario $\beta_{\text{parenting} \rightarrow \text{child outcome}}$ will likely be an over-estimate of the effect, and statistical significance of $\beta_{\text{parenting} \rightarrow \text{child outcome}}$ may represent a false positive.

7.5.3 Simpson's Paradox

Earlier, in [Section 6.3](#), I referred to the idea of *multilevel structure* wherein failure to account for the clustering of data—omission of u_0 from the equation being estimated—may lead to incorrect conclusions. A closely related phenomenon is that of *Simpson's Paradox* (Simpson, 1951) wherein omission of a relevant *covariate* (e.g. SES, community characteristics, country level characteristics) may also lead to dramatically incorrect results.

Statistically, we imagine a situation where the true model is:

$$\text{child outcome}_{it} = \beta_0 + \beta_1 \text{parenting}_{it} + \beta_2 \text{community or country characteristic}_{it} + e_{it}$$

If *community or country characteristics* in fact influence *outcome*, but are not included in the statistical model, perhaps because they are not measured in the data, then the estimate of β_1 for *parenting* will be biased. See [Figure 7.4](#) for an illustration.

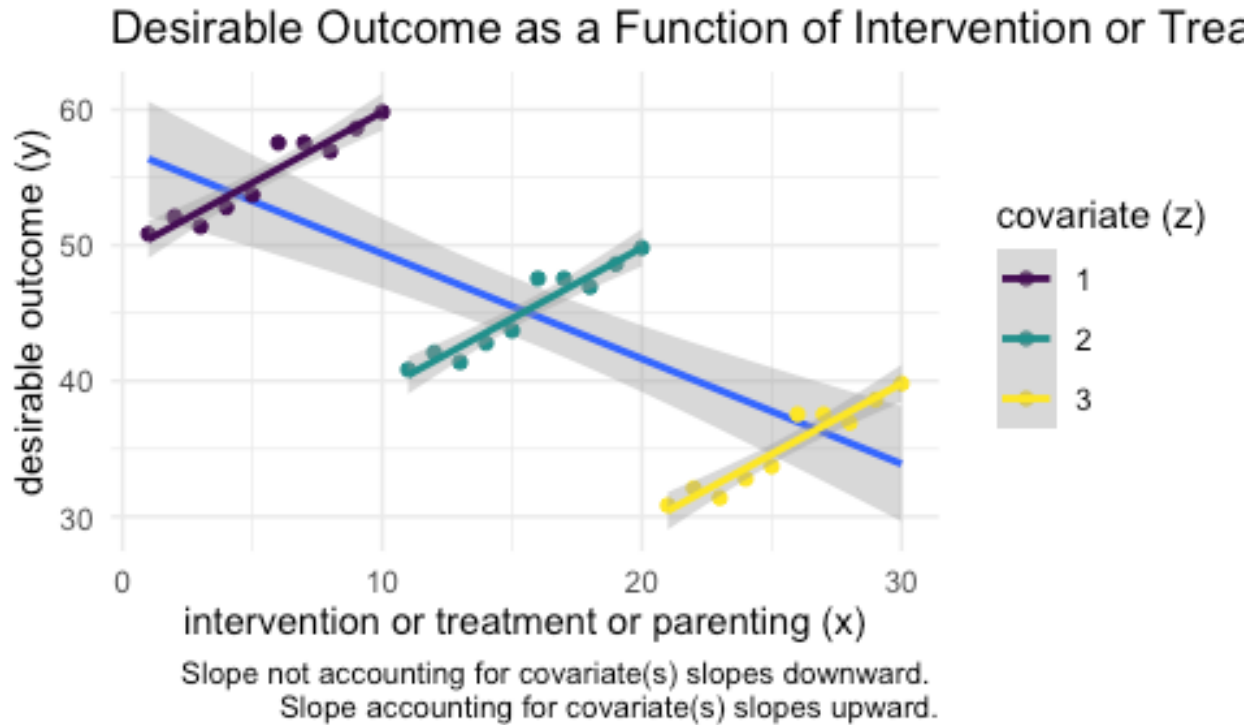


Figure 7.4: An Illustration of Simpson's Paradox

7.5.4 A Simpler Multilevel Model To Explore Causality

For purposes of explication of ideas about causal estimation, in this section, I imagine a simpler equation where I am only considering the clustering of *person timepoints* within *individual people*, and ignoring for the moment—again for the sake of exposition—the clustering of *individuals* within *countries*.

After explication and comprehension of this model, however, it is a simple matter to add back in the random effects for country level clustering.

The appropriate multilevel model is below.

$$\text{outcome}_{it} = \beta_0 + \beta_1 \text{parental warmth}_{it} + \beta_2 \text{physical punishment}_{it} + \beta_3 \text{time}_{it} + v_{0i} + e_{it} \quad (7.2)$$

Note that in [Equation 7.2](#), if one were estimating a *multilevel model*, one would consider the v_{0i} to be a randomly varying parameter with a mean of 0, and a variance of $\sigma^2(v_{0i})$.

7.5.5 Fixed Effects Regression

I can use the same equation:

$$\text{outcome}_{it} = \beta_0 + \beta_1 \text{parental warmth}_{it} + \beta_2 \text{physical punishment}_{it} + \beta_3 \text{time}_{it} + v_{0i} + e_{it} \quad (7.3)$$

However, in [Equation 7.3](#), I now consider the v_{0i} to be *estimable* for each individual i in the data. In effect, the v_{0i} become a unique indicator variable for each individual in the data set. This is known as a *fixed effects regression model*.

Details are provided in Allison (2009) and Wooldridge (2010). StataCorp (2021b) provides an exceptionally clear explication of the core idea of fixed effects regression. The essential idea is that the fixed effects model provides statistical control for all time invariant characteristics of study participants, such as—as is often the case in many data sets—their racial or ethnic identity, or neighborhood of residence, or other characteristics which by definition are time invariant, such as the region of the country or city in which a respondent was born.

Thus, by ruling out many potential confounds, fixed effects regression methods provide much more causally robust analyses, specifically because they control for many more possible confounding variables than do standard regression methods, including multilevel models, which are only able to control for the variables that are measured in the study **and** included within the regression model.

However, a disadvantage of the fixed effects approach is that this approach can not provide estimates for any time invariant characteristic of study participants. Indeed, if one includes time invariant variables into a fixed effects regression, they are automatically dropped from the regression results as can be seen in the regression table below.

The relevant Stata commands are:

- `mixed outcome t warmth physical_punishment || id:`
- `xtreg outcome t warmth physical_punishment, i(id) fe`

	MLM		FE	
t	3.710	*	4.093	*
warmth	4.921	**	4.883	**
physical_punishment	-3.330	**		
Intercept	29.811		-298.721	**
var(_cons)	11955.673			
var(e)	2874.374			
Number of observations			3000	

** p<.01, * p<.05

7.5.6 The Correlated Random Effects Model

The *correlated random effects* model is based upon ideas first developed by Mundlak (1978) and later explicated in Wooldridge (2010). Antonakis et al. (2021) and Schunck (2013) provide very intuitive explanations of this model.

The central idea is that one can obtain estimates of both the time invariant variables, and estimates for time varying variables. The key idea is that for time varying variables, I include the individual level mean for that variable in the model. Thus, in the example below, I include $\beta_{1a}\overline{\text{parental warmth}_i}$.

$$\text{outcome}_{it} = \beta_0 + \beta_1 \text{parental warmth}_{it} + \beta_{1a} \overline{\text{parental warmth}_i} + \beta_2 \text{physical punishment}_{it} + \beta_3 \text{time}_{it} + \quad (7.4)$$

$$v_{0i} + e_{ij}$$

By including this parameter, I obtain estimates for the time varying variables that are *equivalent* to what I would obtain from a fixed effects regression (Schunck, 2013).

The Stata command for the correlated random effects model is:

```
mixed outcome t warmth mean_warmth physical_punishment || id:
```

	MLM		FE		CRE	
t	3.710	*	4.093	*	4.093	*
warmth	4.921	**	4.883	**	4.883	**
physical_punishment	-3.330	**			-3.345	**
mean_warmth					0.295	
Intercept	29.811		-298.721	**	2.219	
var(_cons)	11955.673				11946.798	
var(e)	2874.374				2874.229	
Number of observations			3000			

** p<.01, * p<.05

8. Conclusion

Many data sets relevant to the study of important social issues, or social problems, are inherently multilevel. For example, data on diverse children in schools, diverse individuals in neighborhoods, and individuals or families in diverse and different countries all have multilevel structures in which individuals are clustered in higher level social structures. Data with repeated measures, sometimes termed panel data, can also be thought of as multilevel data sets, wherein individual timepoints are nested inside individuals, who may in turn be nested or clustered in larger social units such as countries. Failure to use appropriate basic multilevel models with such multilevel data can lead to answers that are either biased, or demonstrably wrong. Simple multilevel models allow the researcher to correctly estimate statistical significance, and to correctly estimate regression coefficients while accounting for multilevel structure. More advanced applications of multilevel models allow the researcher to explore the variation in both predictors and outcomes—and the relationship of predictors to outcomes—and to characterize the extent of this variation. Lastly, multilevel models provide a foundation for thinking about closely related models—fixed effects regression, and correlated random effects models—that provide methods for estimation that afford stronger causal conclusions. Thus, for applied researchers, interested in addressing a variety of social problems and social issues with diverse samples of individuals, multilevel models present a method to think clearly about variation, to explore that variation, and to extend that thinking about variation to estimate more causally robust models within the context of diversity and variation.

References

Abelson, R. P. (1995). Statistics as principled argument. In *Statistics as principled argument*. (pp. 221, xv, 221–xv). Lawrence Erlbaum Associates, Inc.

- Allison, P. (2009). *Fixed effects regression models*. Sage Publishing.
- Antonakis, J., Bastardo, N., & Ronko, M. (2021). On ignoring the random effects assumption in multilevel models: Review, critique, and recommendations. *Organizational Research Methods*, 24, 443–483. <https://doi.org/10.1177/1094428119877457>
- Aron, A., & Corne, S. (1994). Introduction. In A. Aron & S. Corne (Eds.), *Writings for a liberation psychology*. Harvard University Press.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Barth, R. P., & Olsen, A. N. (2020). Are children oppressed? The timely importance of answering this question. *Children and Youth Services Review*, 110, 104780. <https://doi.org/10.1016/j.childyouth.2020.104780>
- Bland, J. M., & Altman, D. G. (1994). Statistics notes: Correlation, regression, and repeated data. *BMJ*, 308, 896. <https://doi.org/10.1136/bmj.308.6933.896>
- Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., & Majid, A. (2022). Over-reliance on English hinders cognitive science. *Trends in Cognitive Sciences*. <https://doi.org/10.1016/j.tics.2022.09.015>
- Burkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the r package brms. *The R Journal*, 10(1), 395–411.
- Burton, M., & Kagan, C. (2005). Liberation Social Psychology: Learning From Latin America Psychology of liberation: Learning from Latin America. *Journal of Community & Applied Social Psychology*, 15. <https://doi.org/10.1002/casp.786>
- Cesaire, A. (1956). *Letter to Maurice Thorez*.
- Cokley, K., & Awad, G. (2013). In defense of quantitative methods: Using the “master’s tools” to promote social justice. *Journal for Social Action in Counseling & Psychology*, 5, 26. <https://doi.org/10.33043/JSACP.5.2.26-41>
- Deater-Deckard, K., Dodge, K. A., Bates, J. E., & Pettit, G. S. (1996). Physical discipline among African American and European American mothers: Links to children’s externalizing behaviors. *Developmental Psychology*, 32(6), 1065–1072. <https://doi.org/10.1037/0012-1649.32.6.1065>
- Diener, E., Northcott, R., Zyphur, M. J., & West, S. G. (2022). Beyond experiments. *Perspectives on Psychological Science*, 17, 1101–1119. <https://doi.org/10.1177/17456916211037670>
- Draper, C. E., Barnett, L. M., Cook, C. J., Cuartas, J. A., Howard, S. J., McCoy, D. C., Merkley, R., Molano, A., Maldonado-Carreño, C., Obradović, J., Scerif, G., Valentini, N. C., Venetsanou, F., & Yousafzai, A. K. (2022). Publishing child development research from around the world: An unfair playing field resulting in most of the world’s child population under-represented in research. *Infant and Child Development*, n/a, e2375. <https://doi.org/10.1002/icd.2375>
- Duncan, G. J., & Gibson-Davis, C. M. (2006). Connecting Child Care Quality to Child Outcomes: Drawing Policy Lessons from Nonexperimental Data. *Evaluation Review*, 30(5), 611–630. <https://doi.org/10.1177/0193841X06291530>

- Eamon, M. K. (2001). Poverty, Parenting, Peer, and Neighborhood Influences on Young Adolescent Antisocial Behavior. *Journal of Social Service Research*, 28(1), 1–23.
https://doi.org/10.1300/J079v28n01_01
- Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., Koss, M. P., & Marks, J. S. (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults: The adverse childhood experiences (ACE) study. *American Journal of Preventive Medicine*, 14(4), 245–258.
[https://doi.org/https://doi.org/10.1016/S0749-3797\(98\)00017-8](https://doi.org/https://doi.org/10.1016/S0749-3797(98)00017-8)
- Firebaugh, G. (2001). Ecological fallacy, statistics of. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social & behavioral sciences* (pp. 4023–4026). Pergamon.
<https://doi.org/https://doi.org/10.1016/B0-08-043076-7/00765-8>
- Frank, M. (2018). *Mixed effects models: Is it time to go Bayesian by default?*
<http://babieslearninglanguage.blogspot.com/2018/02/mixed-effects-models-is-it-time-to-go.html>
- Gelman, A., Shor, B., Bafumi, J., & Park, D. (2007). Rich state, poor state, red state, blue state: What's the matter with connecticut? *Quarterly Journal of Political Science*, 2, 345–367.
<https://doi.org/10.2139/ssrn.1010426>
- Gershoff, E. T., & Grogan-Kaylor, A. (2016a). Race as a Moderator of Associations Between Spanking and Child Outcomes. *Family Relations*, 65(3), 490–501.
<https://doi.org/10.1111/fare.12205>
- Gershoff, E. T., & Grogan-Kaylor, A. (2016b). Spanking and child outcomes: Old controversies and new meta-analyses. *Journal of Family Psychology*, 30, 453–469.
<https://doi.org/10.1037/fam0000191>
- Gershoff, E. T., Grogan-Kaylor, A., Lansford, J. E., Chang, L., Zelli, A., Deater-Deckard, K., & Dodge, K. A. (2010). Parent discipline practices in an international sample: Associations with child behaviors and moderation by perceived normativeness. *Child Development*, 81, 487–502.
<https://doi.org/10.1111/j.1467-8624.2009.01409.x>
- Grogan-Kaylor, A., Castillo, B., Pace, G. T., Ward, K. P., Ma, J., Lee, S. J., & Knauer, H. (2021). Global perspectives on physical and nonphysical discipline: A Bayesian multilevel analysis. *International Journal of Behavioral Development*. <https://doi.org/10.1177/0165025420981642>
- Grogan-Kaylor, A., Ma, J., Lee, S. J., Castillo, B., Ward, K. P., & Klein, S. (2018). Using Bayesian analysis to examine associations between spanking and child externalizing behavior across race and ethnic groups. *Child Abuse and Neglect*, 86, 257–266.
<https://doi.org/10.1016/j.chiabu.2018.10.009>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). *The weirdest people in the world?*
<https://doi.org/10.1017/S0140525X0999152X>
- Hines, C. T., Kalil, A., & Ryan, R. M. (2022). Differences in Parents' Attitudes Toward Spanking Across Socioeconomic Status and Region, 1986–2016. *Social Indicators Research*, 160(1), 133–158. <https://doi.org/10.1007/s11205-021-02803-7>
- Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, 81(396), 945–960. <https://doi.org/10.1080/01621459.1986.10478354>

- Hooper, R. (2022). Designing Stepped Wedge Trials with Continuous Recruitment. *Methods: Mind the Gap (Webinar Series)*.
- Khaleque, A., & Rohner, R. P. (2002). Perceived parental acceptance-rejection and psychological adjustment: A meta-analysis of cross-cultural and intracultural studies. *Journal of Marriage and Family*, 64, 54–64. <https://doi.org/10.1111/j.1741-3737.2002.00054.x>
- Kottak, C. (2021). *Anthropology: Appreciating human diversity* (19th ed.). McGraw Hill.
- Kreft, I., & de Leeuw, J. (1998). *Introducing multilevel modeling*. SAGE Publications. <https://doi.org/10.4135/9781849209366>
- Lee, S. J., Ward, K. P., Grogan-Kaylor, A., & Singh, V. (2022). Anxiety and depression during COVID-19: Are adults in households with children faring worse? *Journal of General Internal Medicine*, 37, 1328–1330. <https://doi.org/10.1007/s11606-021-07256-9>
- Luke, D. (2004). *Multilevel modeling*. SAGE Publications, Inc. <https://doi.org/10.4135/9781412985147>
- Ma, J., Grogan-Kaylor, A. C., Pace, G. T., Ward, K. P., & Lee, S. J. (2022). The association between spanking and physical abuse of young children in 56 low- and middle-income countries. *Child Abuse & Neglect*, 129, 105662. <https://doi.org/10.1016/j.chiabu.2022.105662>
- Martin-Baro, I. (1994). Public opinion research. In A. Aron & S. Corne (Eds.), *Writings for a liberation psychology*. Harvard University Press.
- Martin-Baro, I. (1998). Retos y perspectivas de la psicología latinoamericana. In A. Blanco (Ed.), *Psicología de la liberación*. Trotta.
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language*. <https://doi.org/10.1016/j.jml.2017.01.001>
- Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46, 69–85. <https://doi.org/10.2307/1913646>
- Nalborczyk, L., Batailler, C., Loevenbruck, H., Vilain, A., & Burkner, P.-C. (2019, May). An introduction to Bayesian multilevel models using brms: A case study of gender effects on vowel variability in standard Indonesian. PsyArXiv. https://doi.org/10.1044/2018_JSLHR-S-18-0006
- Oberauer, K. (2022). The Importance of Random Slopes in Mixed Models for Bayesian Hypothesis Testing. *Psychological Science*. <https://doi.org/doi:10.1177/09567976211046884>
- Pace, G. T., Lee, S. J., & Grogan-Kaylor, A. (2019). Spanking and young children's socioemotional development in low- and middle-income countries. *Child Abuse and Neglect*, 88, 84–95. <https://doi.org/10.1016/j.chiabu.2018.11.003>
- Poincare, H. (1908). *Science et methode*. Flammarion.
- Proctor, R. N. (2012). The history of the discovery of the cigarette–lung cancer link: evidentiary traditions, corporate denial, global toll. *Tobacco Control*, 21(2), 87 LP–91. <https://doi.org/10.1136/tobaccocontrol-2011-050338>

- Rabe-Hesketh, S., & Skrondal, A. (2012). Multilevel and longitudinal modeling using stata - volume i: Continuous responses. In *Stata Press* (p. 974).
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (pp. xxiv, 485 p.). Sage Publications.
- Rothenberg, W. A., Ali, S., Rohner, R. P., Lansford, J. E., Britner, P. A., Giunta, L. D., Dodge, K. A., Malone, P. S., Oburu, P., Pastorelli, C., Skinner, A. T., Sorbring, E., Steinberg, L., Tapanya, S., Tirado, L. M. U., Yotanyamaneewong, S., Alampay, L. P., Al-Hassan, S. M., Bacchini, D., ... Deater-Deckard, K. (2022). Effects of parental acceptance-rejection on children's internalizing and externalizing behaviors: A longitudinal, multicultural study. *Journal of Child and Family Studies*, 31, 29–47. <https://doi.org/10.1007/s10826-021-02072-5>
- Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology*, 20, 416–420. <https://doi.org/10.1093/beheco/arn145>
- Schunck, R. (2013). Within and between estimates in random-effects models: Advantages and drawbacks of correlated random effects and hybrid models. *The Stata Journal*, 13, 65–76. <https://doi.org/10.1177/1536867X1301300105>
- Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 13, 238–241. <http://www.jstor.org/stable/2984065>
- Singer, J. D., & Willett, J. B. (2003). Applied longitudinal data analysis : Modeling change and event occurrence. In *Applied longitudinal data analysis : modeling change and event occurrence*. Oxford University Press.
- Stage, F. K., & Wells, R. S. (2014). Critical quantitative inquiry in context. *New Directions for Institutional Research*, 2013, 1–7. <https://doi.org/10.1002/ir.20041>
- StataCorp. (2021a). *Stata 17 longitudinal data/panel data reference manual*. Stata Press.
- StataCorp. (2021b). *Stata statistical software: Release 17*. StataCorp LLC.
- Su, F. E. (2017). Mathematics for human flourishing. *The American Mathematical Monthly*, 124, 483–493. <https://doi.org/10.4169/amer.math.monthly.124.6.483>
- UNICEF. (2014). *Hidden in plain sign: A statistical analysis of violence against children* (pp. 1–206). UNICEF.
- UNICEF. (2021). *Multiple indicator cluster surveys (MICS)*. UNICEF. <https://mics.unicef.org/>
- United Nations. (2022). *Sustainable development goals*.
- United Nations General Assembly. (1989). *Convention on the rights of the child*.
- Waddington, H. S., Villar, P. F., & Valentine, J. C. (2022). Can non-randomised studies of interventions provide unbiased effect estimates? A systematic review of internal replication studies. *Evaluation Review*, 0193841X221116721. <https://doi.org/10.1177/0193841X221116721>
- Ward, K. P., Grogan-Kaylor, A. C., Pace, G. T., Cuartas, J., & Lee, S. J. (2021). A Multilevel Ecological Analysis of the Predictors of Spanking Across 65 Countries. *BMJ Open*, 11(e046075). <https://doi.org/10.1136/bmjopen-2020-046075>

Ward, K. P., Grogan-Kaylor, A., Ma, J., Pace, G. T., & Lee, S. J. (2021). *Associations between 11 parental discipline behaviors and child outcomes across 60 countries*. PsyArXiv. <https://doi.org/10.31234/osf.io/f5t8x>

Ward, K. P., Lee, S. J., Grogan-Kaylor, A. C., Ma, J., & Pace, G. T. (2022). Patterns of Caregiver Aggressive and Nonaggressive Discipline Toward Young Children in Low- and Middle-Income Countries: A Latent Class Approach. *Child Abuse & Neglect*, 128. <https://doi.org/https://doi.org/10.1016/j.chiabu.2022.105606>

Watts, D. J. (2011). *Everything is obvious: *Once you know the answer*. Crown Business.

Wiest, L. R., Higgins, H. J., & Frost, J. H. (2007). Quantitative literacy for social justice. *Equity & Excellence in Education*, 40, 47–55. <https://doi.org/10.1080/10665680601079894>

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. The MIT Press. <http://www.jstor.org/stable/j.ctt5hhcfr>