Multilevel Thinking

Discovering Diversity, Universals, and Particulars in Cross-Cultural Research

Andrew Grogan-Kaylor

2024-04-06

Table of contents

List of Figures

List of Tables

1 The Usefulness of Multilevel Modeling and Multilevel Thinking

"I am because we are; and since we are, therefore I am." (Mbiti, 1970)

For decades now, multilevel models have been an important quantitative tool for social research. While multilevel models have become very common in social research, there are aspects of these models that are explored less frequently in published articles that appear in academic journals. This book 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.

The book started out as a set of notes on *things I only get to discuss during breaks, or after class, or during office hours* in my class on *Multilevel and Longitudinal Modeling*, and has grown from that set of notes into an introduction to multilevel modeling.

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.

I start by presenting some initial ideas about multilevel modeling. First, as is relatively commonly understood, *multilevel models allow for the correct estimation of p values in the presence of data clustering*. Second, as is less commonly appreciated, when data are clustered, *multilevel models correctly estimate β regression coefficients and may avoid estimating a regression coefficient that is too large, too small, or even has the wrong sign.*

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 two ideas: *multilevel models as the exploration of diversity and variation across countries and cultures*; and *multilevel models as a foundation for models that let us think more rigorously about causality.* I argue that multilevel models provide a foundation for engaging with cross-cultural diversity in a quantitatively rigorous fashion.

Certainly, none of the statistical ideas contained in this book are unique to me. There are thorough–and often much more mathematically rigorous–presentations of many of the ideas contained in this book 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 (2022)'s more recent and extremely comprehensive two volume text. Luke (2004), and Kreft & de Leeuw (1998), offer shorter, less mathematically rigorous, but still excellent introductions to the topic of multilevel modeling. Gelman et al. (2007) introduced me to the ideas that in this book I describe as "multilevel structure" using an example with voting patterns.

My intent in this book 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 mathematically rigorous 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 Julie Ma, Dr. Kaitlin Ward, Professor Garrett Pace and Professor Shawna Lee. I'm thankful for their collegiality, their friendship, their patience with me, and their dedication to good science. Working with these colleagues has greatly deepened my understanding of multilevel models. 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 book, and probing thoughtful questions. Shari Grogan-Kaylor had many important and challenging questions about the *what* and the *why* of this book, and most importantly to me, *has believed* in it. Don Deutsch showed ongoing interest in the development of this book, and has asked some hard questions that have improved its logic. Thank you to Professor Jim Allen for our discussions about how quantitative methods can better listen to lived experience. 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 Yoonsun Han, Professor Berenice Castillo, Professor Maria Galano, Dr. Sara Stein, Madhur Singh, and Tong Suo. Thank you for the many valuable discussions during class breaks, and after class. These discussions have also greatly extended my understanding of multilevel models. Lastly, I'd like to thank Marty Betts and Greg Knollmeyer for their help and support. While I'm thankful for the inspiration and colleagueship provided by others, any remaining errors and omissions in this book are of course my responsibility.

License and Citation

License

Citation

For attribution, please cite as:

Grogan-Kaylor (2023). *Multilevel Thinking*. Retrieved from https://agrogan1.github.io/multilevel-thinking

BibTeX Citation

```
@book{,
   author = {Andrew Grogan-Kaylor},
   city = {Ann Arbor, MI},
   title = {Multilevel Thinking},
   url = {https://agrogan1.github.io/multilevel-thinking/},
   year = {2023},
}
```

Some Preliminary Thoughts

> “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.”

— (Dalton, 2000) (translated 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 disenthrall ourselves, bestow ourselves to silence, or a severer listening …”

— (Rich, 1984)

> “Research is formalized curiosity. It is poking and prying with a purpose.”

— (Hurston, 1942)

2 Introduction

> "Sure, it's hard to get started; remember learning to use knife and fork? Dig in: you'll never reach bottom. It's not like it's the end of the world–just the world as you think you know it." (Dove, 1999)

> "Listening to the world. Well, I did that, and I still do it. I still do it." (Oliver & Tippett, 2015)

2.1 Quantitative Methods and Social Justice

There is clearly need for both qualitative and quantitative methods. Central to the argument of this book 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. First, 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). I note that one of the pioneers of liberation psychology, Martin-Baró (Aron & Corne, 1994), used both qualitative and quantitative methods (Martin-Baro, 1994a), including in the latter case, relatively sophisticated arguments about patterns of missing data across a survey data set (Aron & Corne, 1994).

Second, when we have samples of a hundred, several hundred, several thousand, or even hundreds of thousands of study participants distributed across multiple and diverse social contexts, it is difficult to imagine a methodology other than a multilevel 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 *diversity and 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.

Therefore, I consider multilevel modeling to be a principled quantitative method for listening to the voices of large numbers of study participants across social contexts. In Section **??**, where I consider the estimation of p values, and Section **??**, where I consider the signs of regression coefficients, I explore the ways that *multilevel data* can contribute substantially to the complexity of the analysis of data. I thus argue that advanced quantitative methods, like multilevel modeling, can play an important role in contributing to liberatory ideas.

There is thus an ethical argument that is embedded in this book. 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 polices, 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 book is that a deeper study of multilevel modeling can result in an advanced "quantitative literacy" (Wiest et al., 2007), "quantitative criticalism" (Scharrer & Ramasubramanian, 2021), or "principled argument" (Abelson, 1995), that is appropriate for drawing accurate conclusions from multilevel data.

2.2 Some Philosophy of Science

I am not much of a philosopher of science. However, I am very persuaded by Strevens' (2020) minimalist criterion of the "iron rule". In essence, this rule specifies that to count as "science", investigations must engage in "performing an experiment or making an observation that generates relevant empirical evidence" against which competing hypotheses can be tested. A similar perspective is offered by Goldacre (2011) who argues that ideas about interventions should be scrutinized with a "fair test". That is to say, they should be tested against evidence that can support or refute those ideas. I would argue all ideas about promoting human well-being should be able to be subjected to such a "fair test".

I believe that our work—whether qualitative, or quantitative—should strive to be both critical *and* scientific, in the sense that: our research should gather evidence; that evidence should be assessed in order to support, refute, or modify our initial beliefs; and that evidence should be used to think critically about human wellbeing, including dynamics of power and privilege and disparities. With regard to this idea, Shrader-Frechette (2014) suggests that a "practical philosophy of science" can contribute both to "speaking truth to power" and to "seeking justice".

2.3 A Pragmatic Approach

This book 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. The book takes a very pragmatic and very advocacy oriented approach to improving research.

> "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)).

> "What we see and how we see is of course determined by our perspective, by the place from which we begin our examination of history; but it is determined also by reality itself." (Martin-Baro, 1994b)

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

The book 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 book 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.

2.4 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, or in contexts that are disadvantaged. 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 **??**, Section **??** and Section **??**, I provide specific examples of how multilevel data provides even more opportunity to present answers that are *not* obvious.

2.5 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, Pace, et al., 2021; Ward et al., 2022; Ward et al., 2023). For the presentation in this book, 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*.

2.6 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?

2.7 Universalism And Particularity

> "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)

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, parenting practices may vary widely between cultural groups (Gottlieb, 2002). Further, 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, 2022; 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 social life across countries.

3 Simulated Multi-Country (Multilevel) Data

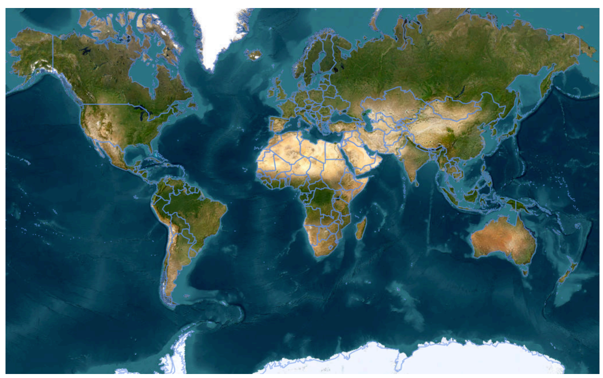

Figure 3.1: Countries of the World

> "… the particular and the universal are not to be seen as opposites, … the universal is not the negation of the particular but is reached by a deeper exploration of the particular." (Cesaire in UNESCO (1997))

I use simulated data in this example. Data come from 30 hypothetical countries. Data contain measures of a few key aspects of parenting[1] or caregiving that have proven salient in the empirical literature on parenting to date: parental `warmth`, and `physical punishment`. Both

[1] I use the term parenting throughout this book, but am aware that such parenting may come from biological parents, or from other caregivers.

parenting measures are normally distributed variables, and are considered to be *Level 1*, or *individual level* variables. `group` is a hypothetical–and somewhat arbitrary–group designation that could hypothetically refer to something like different economic groups, or groups from different parts of the country. `group` is also a *Level 1* variable.

`HDI` is a measure of the *Human Development Index* (United Nations Development Program, 2022), and is measured at the *country level*, or *Level 2*. (I discuss more in depth thinking about levels of the data in Chapter **??**.)

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.

i Download The Data in Stata Format

- Cross-Sectional Data
- Longitudinal Data

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	group	HDI	outcome
1.1	1	3	2	2	69	59.18
1.2	1	0	4	2	69	61.54

Table 3.1: Simulated Multilevel Data

id	country	warmth	physical_punishment	group	HDI	outcome
1.3	1	4	4	1	69	51.87
1.4	1	6	0	2	69	51.71
1.5	1	2	3	2	69	55.88
1.6	1	3	5	1	69	60.78

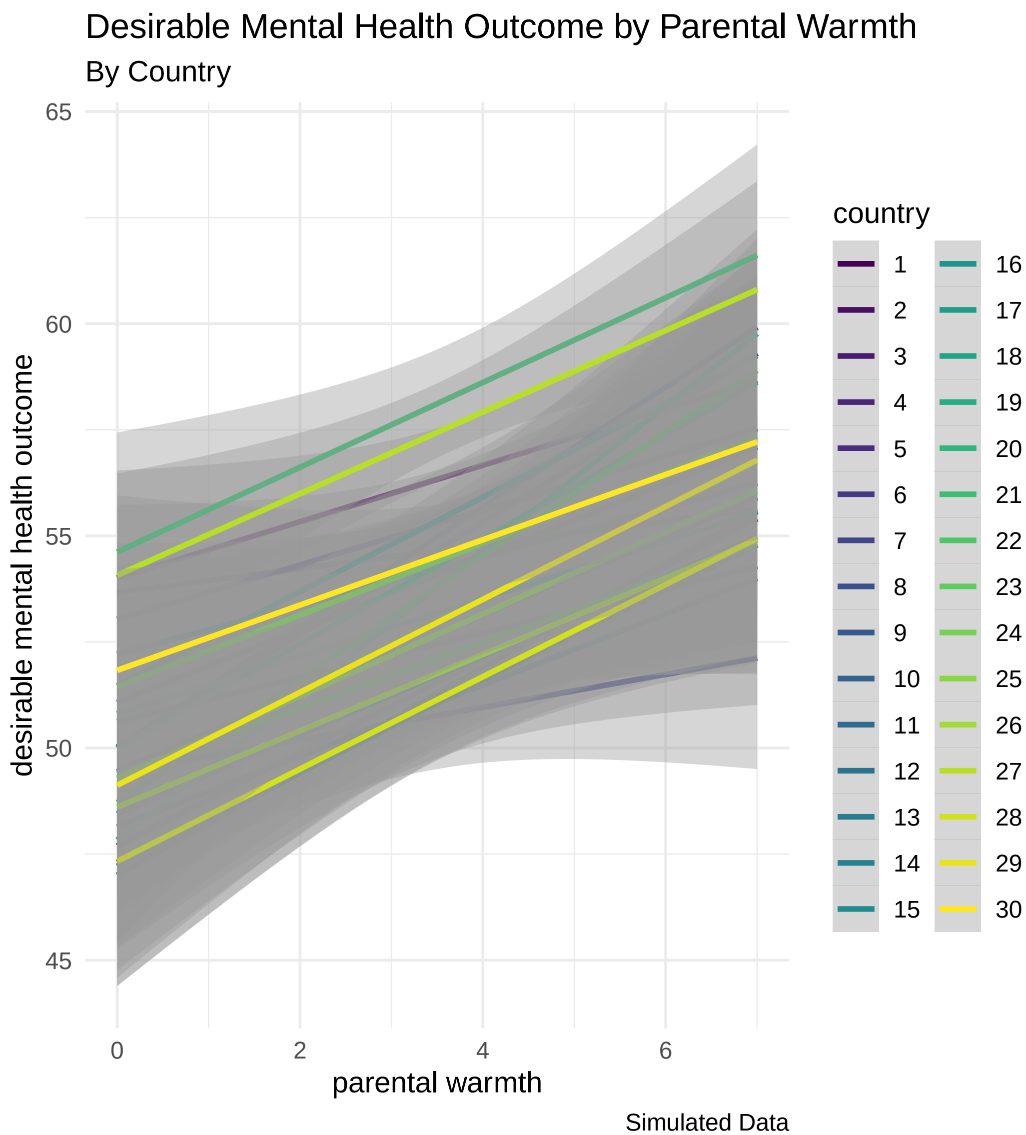

Figure 3.2: Graph of Simulated Data

4 Software

> "– And so, why can't my numbers be beautiful to me? Why the scorn, the doubt in your face? Do you think I am brittle and dusty as old paper? Look again. See the numbers shine in my eyes." (Pye, 2011)

In this book, I use Stata (StataCorp, 2021c) to analyze data. Stata is my software of choice in this book 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	`use mydata.dta`
Descriptive statistics	`summarize x y`
Frequencies	`tabulate x`
Correlation	`corr x y`
Regression	`regress y x`
Logistic Regression	`logit y x, or` [1]
Multilevel Model	`mixed y x || group: x`

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

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

5 Conceptual Framework

> "Ubuntu" defined as: "A person is a person through other people." e.g. in (Mangharam, 2017)

> "The language we have in that world is not large enough for the territory that we've already entered." (Whyte & Tippett, 2016)

5.1 Units of Analysis and Processes at Multiple Levels

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 implications of different ideas about the units of analysis in Section **??**.

Further, with multilevel models, we are not only able to consider the idea of units of analysis at multiple levels of the data, but to consider how variables at both Level 2 and Level 1 may affect an individual level (Level 1) outcome.

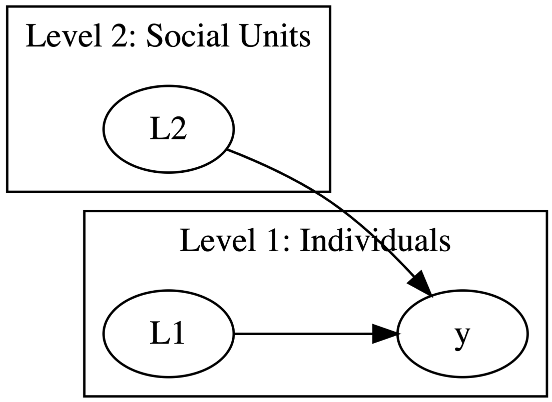

Figure 5.1: Conceptual Framework

5.2 Variables at Multiple Levels

In this book, I distinguish between *conceptual* and *statistical* levels of variables.

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.

Table 5.1: Multiple Levels of Variables

	statistical level 1	statistical level 2
conceptual level 1	Individual response about parenting or mental health	Aggregated responses about parenting or mental health
conceptual level 2	Individual response about community	Aggregated response about community
conceptual level 2	N/A	Administrative indicator of social unit

- Thus, 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.

  Such aggregated variables represent the average level of a response across each Level 2 unit. I could create such a level 2 variable for a variable `x`, using the command: `bysort group: egen mean_x = mean(x)`. For example, in the data described in Chapter **??** I could create a mean country level warmth score with the command `bysort country: egen mean_warmth = mean(warmth)`.

- Using my terminology, community collective efficacy$_{ij}$ or community safety$_{ij}$ would be considered to be a variable that was *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* refer to Level 2 concepts that 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 easily definable 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.

5.3 Multilevel Models As The Exploration Of Variation and Diversity

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 and diversity 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.

5.3.1 A First Example: A Study Of Parenting And Child Development

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*.

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

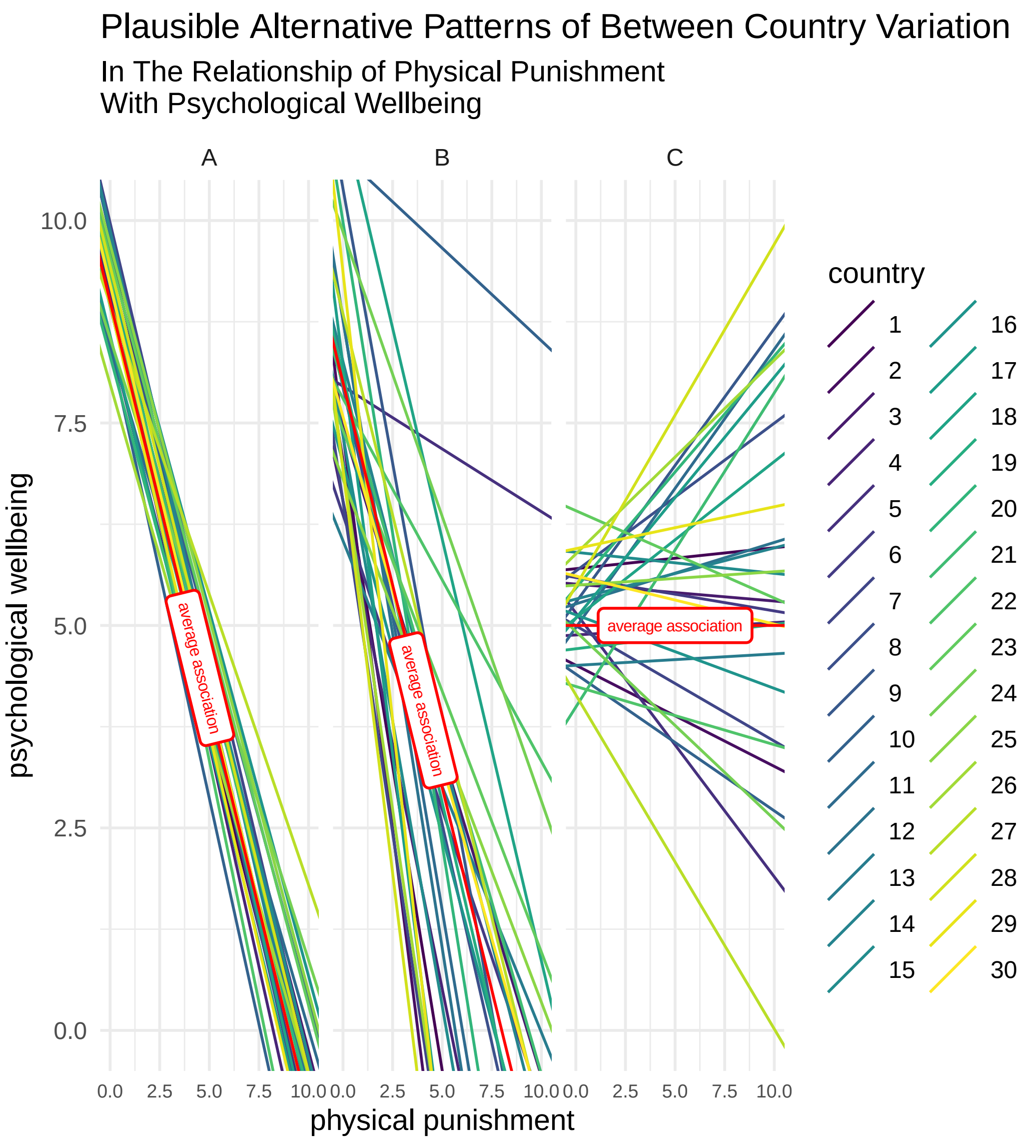

Figure 5.2: Plausible Alternative Patterns of Between Country Variation

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.

5.3.2 A Second Example: A Study Of A Treatment Or Intervention

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.

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.

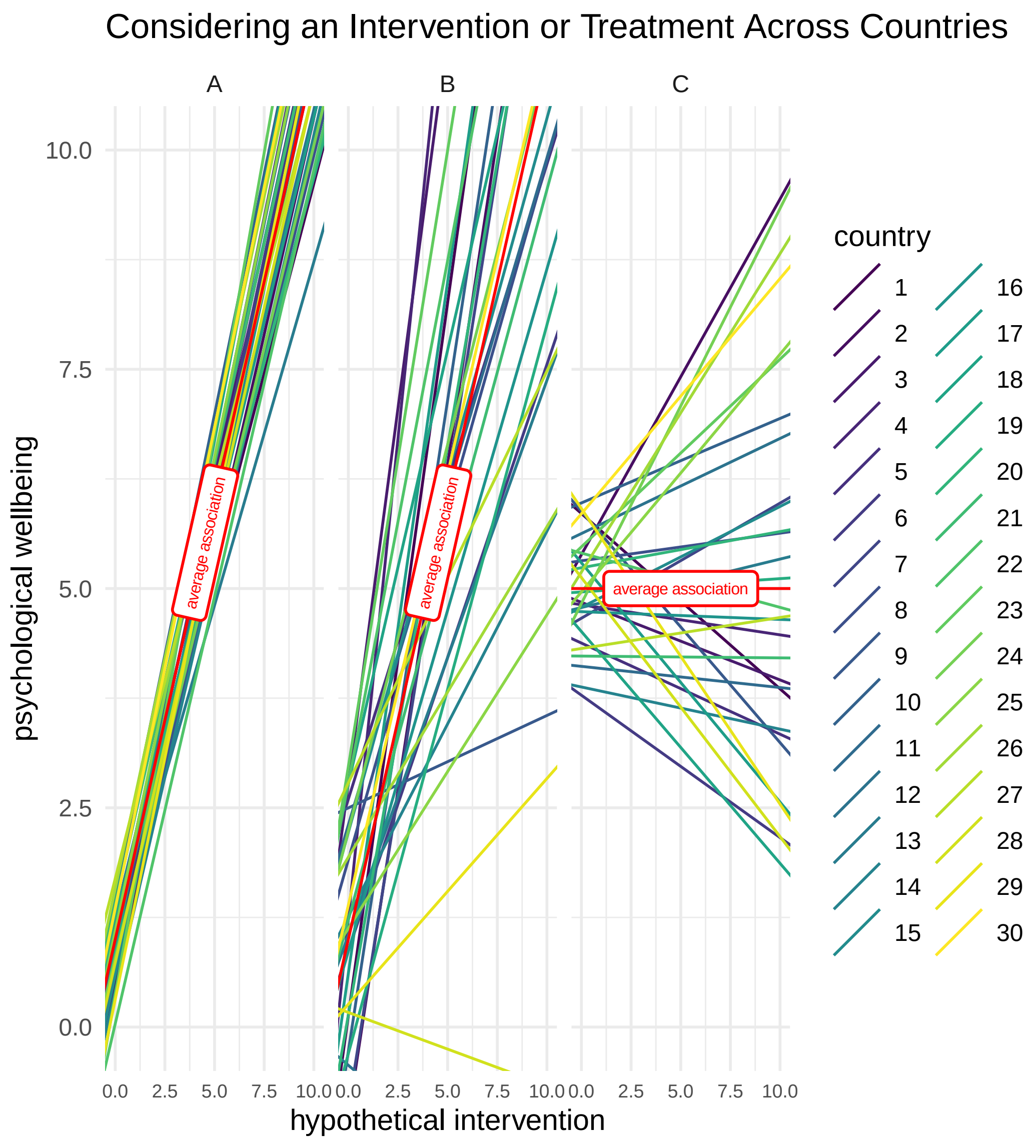

Figure 5.3: Considering an Intervention or Treatment Across Countries

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.

5.3.3 Exploring Variation

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)

As I discuss these ideas in more statistical depth, later in the book, I develop more statistically based ideas about the study of diversity and variation in Section **??**, Section **??**, and Section **??**.

Again, statistically 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, and to concerns of variation and diversity, more clear.

6 The Cross Sectional Multilevel Model

> "Mathematical Science shows us what is. It is the language of unseen relations between things. But to use & apply that language we must be able fully to appreciate, to feel, to seize, the unseen, the unconscious. Imagination too shows us what is, the is that is beyond the senses." (Lovelace, 1992)

I begin this chapter by introducing two key concepts: multilevel models can improve our estimation of p values; multilevel models can improve our estimation of β coefficients.

In these sections I make some initial use of the Stata syntax for regression `regress y x z`, and the Stata syntax for multilevel models, `mixed y x z || groupid:`.

After introducing these two key concepts of multilevel modeling, I then begin a more in depth exploration of the equations and concepts and statistical syntax of the cross sectional multilevel model.

6.1 Estimating Standard Errors And p Values

6.1.1 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.

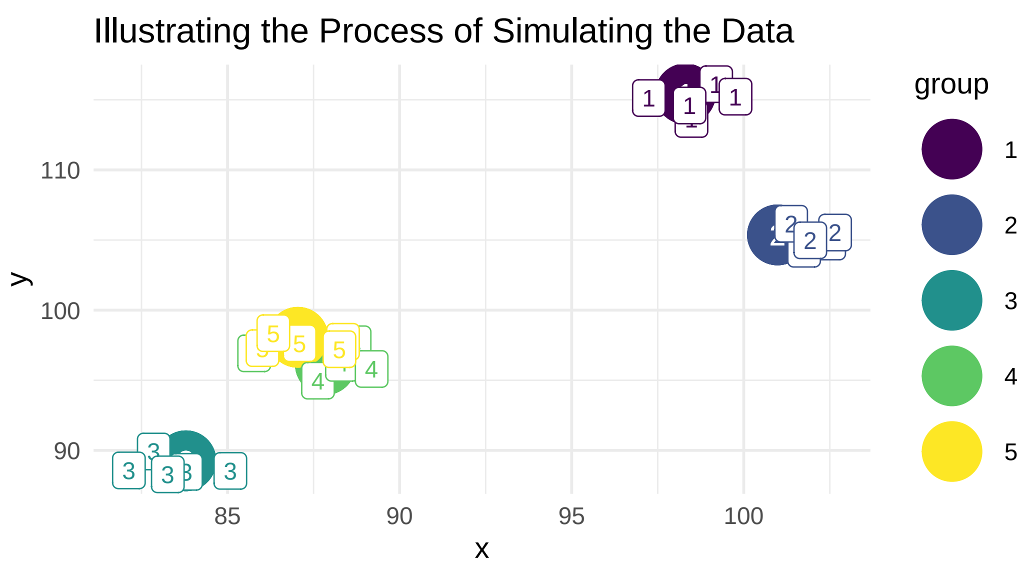

Figure 6.1: Simulated Clustered Data

6.1.2 Compare OLS and MLM

An analysis that is not aware of the grouped nature of the data will give biased results, will misestimate 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:

- OLS: `regress y x`
- Multilevel Model: `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.

6.2 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 teaching 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. ($N_{\text{clusters}} >= 30$)

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.

I use some data simulated for this particular example.

6.2.1 Graphs

6.2.1.1 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.

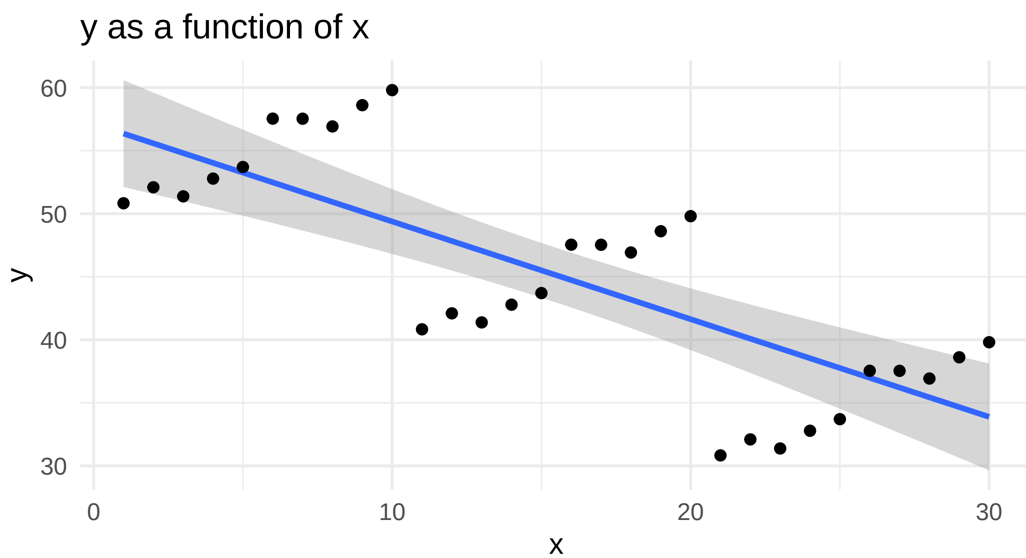

Figure 6.2: A ‘Naive’ Graph

6.2.1.2 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.

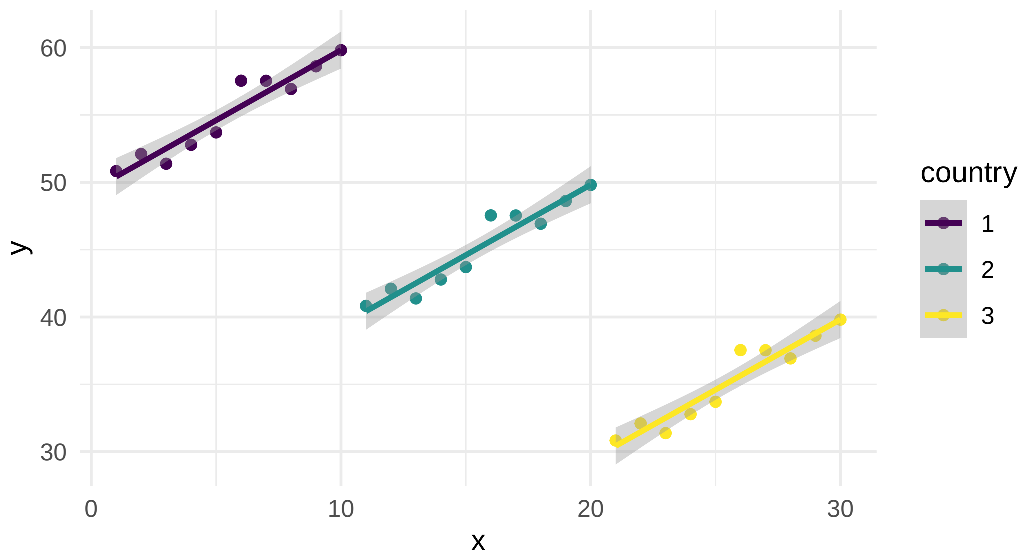

Figure 6.3: An 'Aware' Graph

6.2.2 Regressions

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

The Stata syntax that we use for these analyses is:

- OLS: `regress y x`
- Multilevel Model: `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 reflects this. The coefficient for x in the multilevel model is *positive*, and statistically significant.

	OLS		MLM	
x	-0.775	**	1.038	**
Intercept	57.133	**	29.029	**
var(_cons)			276.867	
var(e)			0.916	
Number of observations	30			

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

6.2.3 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.

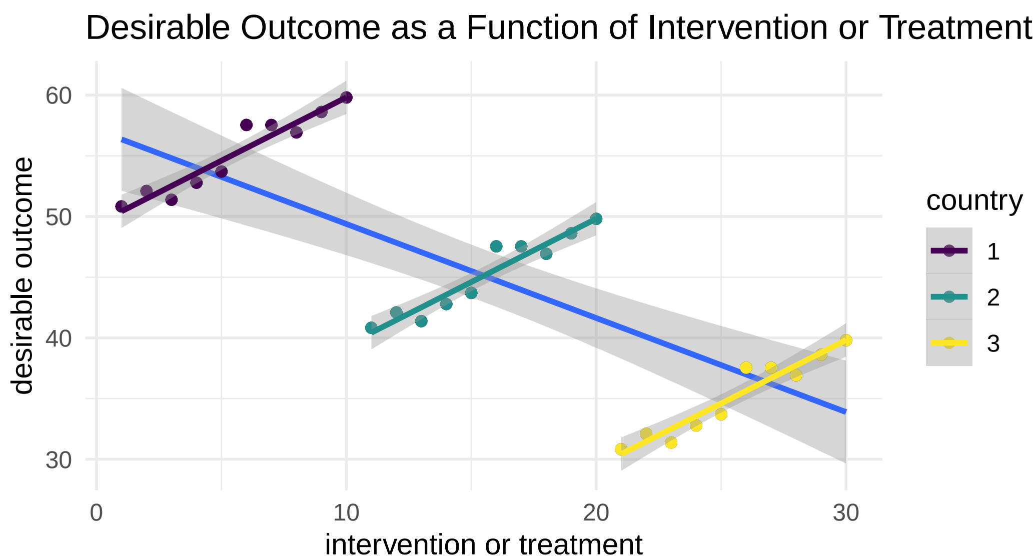

Figure 6.4: A Heuristic Example

The idea that group level and individual level relationships must be the same (Firebaugh, 2001) has been termed the "ecological fallacy".

These data are constructed to provide this kind of extreme example, but it 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.

6.3 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 equivalent.s

6.3.1 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.

6.3.1.1 Level 1 (Individuals)

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + \beta_{2j}z_{ij} + e_{ij} \tag{6.1}$$

6.3.1.2 Level 2 (Countries)

$$\beta_{0j} = \gamma_{00} + \gamma_{01}w_j + u_{0j} \tag{6.2}$$

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

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

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

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's[1] 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. The parameters for the Level 2 equations are represented by γ's. w a Level 2 variable appears in the first Level 2 equation.

[1]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.

6.3.2 One Level of Equations

By simply substituting the values of the Level 2 equations into the Level 1 equations–and rewriting the γ's as β's–we obtain:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 z_{ij} + \beta_3 w_j + u_{0j} + u_{1j} \times x + e_{ij} \tag{6.3}$$

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 w is a Level 2 variable.

> Notice that in this *single equation* format all variables–no matter their *level*–appear in the same equation.

In this formulation of the equation, the nature of the random effects is more clear, and merits discussion. Notice that we have included a *random intercept* u_{0j} as well as a *random slope* $u_{1j} \times x$. The *random intercept*, u_{0j}, indicates that there is variation in the *intercept* of the country specific regression lines, as is true in Figure **??**. The *random slope* term associated with x, $u_{1j} \times x$, indicates that we are allowing for the possibility of variation in the *slope* of the regression lines that is associated with x, in this case, the slope of parental warmth, as is possibly suggested in Figure **??**.

To make these ideas more concrete, I rewrite this equation in terms of the main substantive ideas of this book:

$$\text{outcome}_{ij} = \beta_0 + \beta_1 \text{parental warmth}_{ij} + \beta_2 \text{physical punishment}_{ij} +$$

$$\beta_3 \text{group}_{ij} + \beta_4 \text{HDI}_j +$$

$$u_{0j} + u_{1j} \times \text{parental warmth} + e_{ij} \tag{6.4}$$

Put substantively, this model indicates that the outcome can be conceptualized as a function of an intercept term, and contributions of parental warmth, physical punishment, group membership, and country level HDI. The random intercept, u_{0j} indicates that there is some unexplained variation in the outcome at the country level. The random slope $u_{1j} \times$ parental warmth indicates that the model is allowing for country level variation in the association of parental warmth with the outcome. Inspection of Figure **??** indicates that it might be possible that there would be variation across countries in this slope. The model could be extended to allow for country level variation in other slope terms by adding other random slopes, eg u_{2j}, u_{3j}, etc.

Drawing upon ideas from Chapter **??**, this single level equation can be easily represented in Stata syntax.

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

6.4 Regression With Simulated Multi-Country Data

After considering some of these broader issues, let's now 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.

6.4.1 Unconditional Model

The unconditional model is a model with no x's or covariates (Raudenbush & Bryk, 2002).

$$\text{outcome}_{ij} = \beta_0 + u_{0j} + e_{ij} \tag{6.5}$$

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–attributable to unmeasured variation at the country and individual level.

6.4.2 Intra-Class Correlation Coefficient

I now introduce a measure known as the Intra-Class Correlation Coefficient, (ICC) that can be computed from this unconditional model (Raudenbush & Bryk, 2002).

$$\text{ICC} = \frac{var(u_{0j})}{var(u_{0j}) + var(e_{ij})} \tag{6.6}$$