Stein, R. E., Conley, J. F., and Davis, C. (2016). The differential impact of physical disorder and collective efficacy: A geographically weighted regression on violent crime. *GeoJournal, 81*(3), 351–365.

Waller, L. A., Zhu, L., Gotway, C. A., Gorman, D. M., and Gruenewald, P. J. (2007). Quantifying geographic variations in associations between alcohol distribution and violence: A comparison of geographically weighted regression and spatially varying coefficient models. *Stochastic Environmental Research and Risk Assessment, 21*(5), 573–588.

Wheeler, D. C., and Waller, L. A. (2009). Comparing spatially varying coefficient models: A case study examining violent crime rates and their relationships to alcohol outlets and illegal drug arrests. *Journal of Geographical Systems, 11*(1), 1–22.

TOPIC 8.4 MIXED (A.K.A. MULTILEVEL) MODELS

Ralph B. Taylor

Introduction/Background

You want to read this topic if you have a data set structured in any of the following ways: Do you have cross-sectional observations with data points organized by or clustered by larger spatial units (think surveys of residents clustered around different street corners)? Do you have panel design longitudinal observations (think repeated surveys of the same residents)? Do you have panel design longitudinal observations nested within larger spatial units (think repeated surveys of the same residents in several different neighborhoods)? If you are working with a data set organized in any one of these three ways—grouped into spatial units, temporal units, or both spatial and temporal units—you want to be using multilevel models. Over the past three-plus decades, social scientists have deployed these models in numerous social science disciplines, including political science, public health, criminology, sociology, psychology, and education. This topic includes some details from a worked example. The relevant program and output file appear in Supplement 8.4.

If you want to read research using these models, start with one of the most-cited works (Sampson et al. 1997). This three-level study, nesting items within indexes, and individuals within neighborhoods, explored the impact of collective efficacy (Sampson 2012) on offending and victimization. Other studies on crime looking at temperature effects over time (Sorg and Taylor 2011) or distance from public housing communities (Haberman et al. 2013) may prove of interest. One of the main threads in my past research, often pursued in collaboration with fantastic colleagues, addressed fear of crime (Covington and Taylor 1991), sometimes using different methods (Perkins and Taylor 1996) and sometimes longitudinally (Robinson et al. 2003). Multilevel models provided insights in several works examining the impacts, over a decade, of perceived as well as assessed incivilities on reactions to crime including fear and behavioral avoidance (Taylor 2001). Investigations in related topics, including what drives perceptions of problematic teen groups, a key physical incivility (Taylor et al. 2011); relations between perceptions of gang problems and gang presence (Blasko et al. 2015); and the origins of key elements in collective efficacy (Garcia et al. 2007), benefited from using these models as well. If you are a justice researcher who just happens to have landed in a crime and place handbook, multilevel models of perceptions of police (Taylor et al. 2010) and confidence in the police (Taylor and Lawton 2012) may prove of interest.

What Are They?

Multilevel models are a family of statistical models for analyzing data where those data are clustered in space, in time, or in both (Baumer and Arnio 2012). "Clustered" means that individual observations belong to larger groups of observations, each group of observations sharing some common attribute. The multilevel analysis, when looking at each cluster of observations, simultaneously looks at all the other clusters in the data set, "learns" from the entire data set, and applies that learning to how it "treats" each individual cluster. Of course, clustering can happen at more than one level. For example, you may have surveys of residents nested in different neighborhoods, and those different neighborhoods might themselves be nested within different cities. With clustered data, the data at the lowest level of observation are called level-1 units, and the first grouping units are called level-2 units.

Multilevel models have several different names, including hierarchical linear models, mixed models, random effects models, and random coefficient regression. Do not be confused. They are all in the same family. They also can be applied to a wide range of outcome measures, including binary (used here), nominal, ordinal, count, and, of course, normally distributed outcomes.

These models even work for "imperfect hierarchies" (Snijders and Bosker 2012, p. 205). For example, with a cross-classified model, an observation, such as a person, might belong simultaneously to two different types of units, and both types are at the same level. For example, you may have juveniles adjudicated delinquent who live in different neighborhoods and attend different programs (Lockwood 2011). You can simultaneously consider both aspects of the grouping structure.

The multilevel models also work with multiple membership multiple classification (MMMC) models, "where an observation does not belong simply to one member of a classification" (Browne et al. 2001, p. 103). Here, "a lowest level unit is a member of more than one higher classification unit" (p. 109). For example, especially useful for geography and crime, yearly crime observations might be nested within a neighborhood and that neighborhood might be nested within a neighboring zone of neighborhoods. This could be analyzed using MMMC mixed models.

This text explores only a couple of *basic* features of mixed models. These models can do much more: cross-level interactions, random effects of predictors, moderation at different levels, and so on. These models also can include higher-level predictors, for example, level-2 predictors that are ecological in origin, like census data for a neighborhood or a city, as well as level-1 predictors. If the model has only higher-level predictors, this is called a means as outcomes regression multilevel submodel (Raudenbush and Bryk 2002). For more on any of these points, consult any advanced texts mentioned at the end.

Why Use Them?

The short answer is to get better answers to your questions. Just four quick points: two theoretical, and two analytic. First, applying a monolevel statistical model like multiple regression to clustered data like surveys of residents in different neighborhoods will yield b weights for individual predictors that are potentially confusing. The impact indicated by the b weight may reflect two types of connections between the predictor and the outcome: a connection based on *ecological* covariation at the neighborhood level, and a connection based on *social psychological* covariation, that is, based on predictor-outcome covariation built on differences between neighbors in the same neighborhood, and then pooled across all the neighborhoods. With a monolevel model, you do not know how much the results reflect ecological processes versus social psychological processes. Further, with a monolevel model, statistical tests of b weights may be incorrect because of model error structures (see below).

Second, you might be interested in theoretical connections between the outcome and the predictors at different levels. For example, you might expect that the individual-level connection between resident age and fear of crime varies depending on neighborhood factors (Maxfield 1984). A specific theoretically relevant neighborhood factor, which you might include as a predictor, might influence the impact of an individual factor, like age. This is called a cross-level interaction "because it involves explanatory variables from different levels" (Hox 2010, p. 20). A random coefficient regression multilevel model with an (age × neighborhood variable) interaction would allow you to test this theoretical idea.

Third, conducting a statistical analysis of clustered data while using a monolevel model violates the assumptions underlying that model. For example, if you run something like a multiple regression with clustered data, the resulting error terms probably will be correlated with one another for observations within the same cluster, violating the multiple regression assumption of independent residuals. Incorrect modeling of these error structures can lead to misleading statistical tests of b weights. Naturally, multilevel models make their own assumptions about data structures and sampling procedures. These, of course, can be violated as well. Again, see the advanced texts listed at the end.

Finally, when making estimates for level-2 units, such as, for example, mean neighborhood scores on a fear of crime index, the analysis learns from all of the level-2 neighborhood units to estimate each neighborhood's latent or "true" score on the index. This is one way these models connect up to structural equation modeling (Bauer 2003).

How They Work

The worked example relies on behavioral observations of pedestrians in 24 small commercial centers (SCCs) in Minneapolis and St. Paul completed in the early 1980s (Mcpherson et al. 2006). These 24 SCCs were sampled from a larger set of SCCs in the Twin Cities. More background appears in Supplement 8.4. The binary outcome of interest here (primbiz) is whether the recorded pedestrian seemed to be primarily using a business or service in the center (= 1) or not (= 0). Those scoring 1 are called business-using pedestrians. The predictor of interest (kidrteen) is whether the pedestrian was classified as a child or teen (= 1) or an adult or senior citizen (= 0). Given the binary outcome, a logit model is used.

The "Buckets" Question

The first mixed model has no predictors, just the outcome: the ANOVA (analysis of variance) or null or empty model. This model answers the "buckets" question: How much of the variation in the outcome is in a bucket composed of within-center-between-pedestrian (level-1) variation? How much of the variation is in a bucket composed of between-center (level-2) variation? Further, and more specifically, is the amount of outcome variation in this latter, level-2 bucket, at the level of the SCCs themselves, important? In other words, when predicting business-using pedestrians in different SCCs, do you need to include *multiple constants*, one for each SCC? More simply stated, is ecological variation in the outcome significant?

When you look at the log (output) file in Supplement 8.4 (line # 1172), you will find a likelihood ratio (LR) chi-square (χ^2) test. If it is statistically significant, $p < .05$ typically, it means you need a multilevel logit equation rather than a monolevel logit equation. You should see this:

$$\begin{aligned} &\text{LR test vs. logistic model: chibar2(01)} = 656.33 \\ &\qquad\qquad\qquad\qquad \text{Prob} > = \text{chibar2} = 0.0000 \end{aligned} \tag{8.4.1}$$

Translated, this means the LR χ^2 test with one degree of freedom = 656.33, and there is less than 1 in 10,000 probability that the between-SCC differences in the proportion business-using pedestrians could occur due to chance under the null hypothesis of no differences across SCCs on this outcome. You reject this idea because the test is statistically significant. In other words, you need a multilevel model.

You can learn the size of the level-2 bucket of ecological outcome variation as a proportion of all the outcome variation. After issuing the postestimation command estat icc, you get this proportion, also known as the intraclass correlation, and it is 11.27 percent. This appears in the Supplement 8.4 results under intraclass correlation (line # 1198). The analysis has now separated outcome variation into two parts: level-1, between-pedestrians-within-SCCs; and level-2, between SCCs.

Do not be dismissive of this intraclass correlation because it seems like a small amount of the outcome. *Theoretically* because the LR χ^2 test was statistically significant, *this variation is important* (Liska 1990). Some texts get this wrong and say that if the fraction is small enough, ignore it. Do not believe them. Be guided by whether the LR χ^2 test in your ANOVA model is significant. The practical significance depends on your specific issue being examined.

Estimated "True" Scores on the Outcome at Level-2, the Commercial Center Level

So far, this model with no predictors has told you two things. First, you need a mixed model allowing each commercial center to have its own average proportion on the binary outcome variable. Stated differently, the outcome variation at level-2 is theoretically important. Second, it has told you exactly what portion of the outcome variation is placed at this ecological level. Third, it also can tell you the following: The postestimation predict command generates an estimated "true" proportion of business-using pedestrians for each SCC. These are empirical Bayes estimates. You also can think of them as "latent [hidden] variables" (Snijders and Bosker 2012, p. 62). How are these estimates made?

We do not have time to get into the weeds on Bayesian statistics.[1] The overall idea is that estimates of group means are *shrunken toward* or *biased toward* the overall adjusted mean outcome score. You probably have two questions: Why? How?

The why is as follows. First, the analysis seeks to estimate, for each SCC its *true* mean outcome score, the *true* proportion of business-using pedestrians for that SCC. For each SCC, it is deriving a population estimate, assuming the data records are probability sample data. If you read more about the study (in Appendix A of Supplement 8.4), you will learn that these 24 SCCs were sampled from a larger number of SCCs in Minneapolis-St. Paul. Second, as the analysis seeks to estimate each group's true outcome score, it recognizes that (a) each group proportion is part of a larger set of 24 SCC-level outcome proportions, and (b) the larger set of proportions constellate around an overall, sample-wide proportion of business-using pedestrians. When it asks, for example, where the proportion of business-using pedestrians "should" be for observations at East 15th Street and Nicollet Avenue in Minneapolis, it considers both the center-level proportions in the other 23 centers and the overall proportion.

The how is as follows. In the case of East 15th Street and Nicollet Avenue, the observed business-using pedestrian proportion of .273, the lowest in the set, was empirical Bayes adjusted up to a proportion of .282. That reflects the best guess of the "true" score of the proportion of business-using pedestrians. In making those adjustments, the estimation considers the following. Group outcome means (proportions) are adjusted *more* toward the overall precision weighted outcome mean (proportion) the more each of the following conditions hold for that mean:

1. The original group mean is *farther* from the overall mean.
2. The cases in a group contributing to the group mean *disagree more* with one another on the outcome score, that is, the within group variation on the outcome score is larger.
3. The number of cases in the group is *smaller*.
4. The mean is in a set of group means that, as a set, are *not* widely dispersed.

You can see the original and adjusted proportions in the graph in Figure 8.4.1. Note the following. Each group mean has been adjusted some. Second, all the adjustments are *toward* the adjusted overall average. Those above the mean were adjusted downward; those below the mean were adjusted upward. Third, the more extreme group means were adjusted more, that is, group means farther from the overall mean got shrunken more toward the adjusted overall average mean.

Adding a Predictor

Now you are ready to add the predictor of interest, whether the pedestrian was young, a child or teen (kidrteen = 1), or older (kidrteen = 0). You first look at the raw relationship between these two variables. See the two-way table in Supplement 8.4 (starting line # 878). You see that the younger pedestrians were observed primarily using a business or service only 56.55 percent of the time ($n = 794$), while the *older* pedestrians primarily used a business or service *more frequently*, 70.54 percent of the time ($n = 4{,}015$). This relationship is more than chance ($p < .001$ by LR χ^2 test). Younger pedestrians were less likely to be business-using pedestrians.

This negative relationship, however, blends what are, potentially, two different relationships. There could be one ecological relationship, a between-SCC relationship at the level of the SCCs themselves, and, a second, social psychological, between-pedestrians-within-SCCs relationship. The former would capture level-2 predictor-outcome covariation, and the latter would capture level-1 predictor-outcome covariation. Is it possible the relationship is different at the different levels?

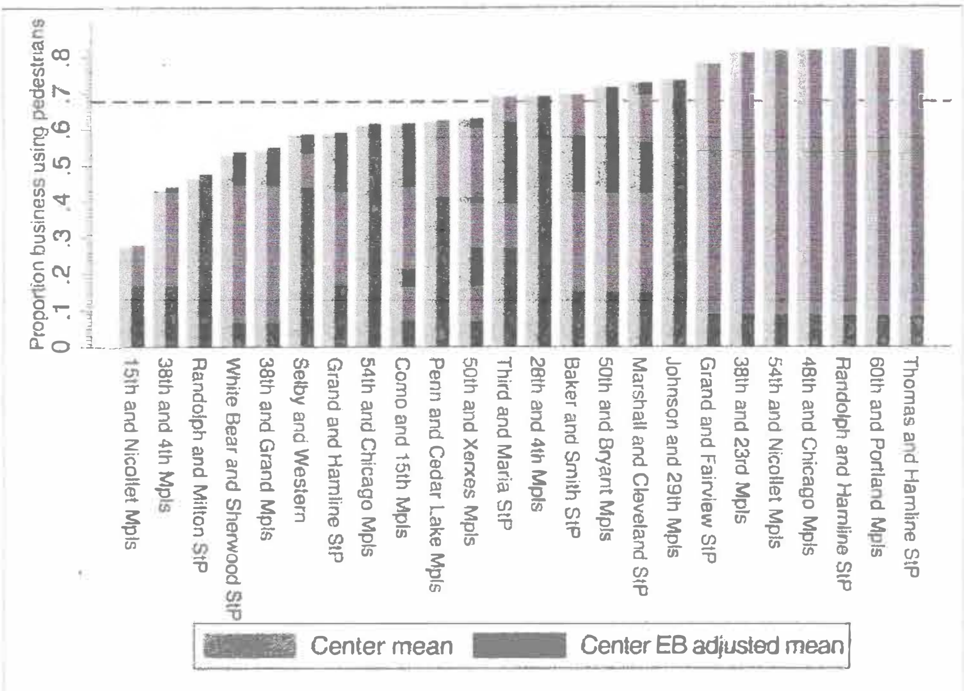

Figure 8.4.1 Observed, and empirically Bayes adjusted, small commercial center-level means for the variable primarily business- or service-using pedestrians (= 1) or not (= 0). *(Author calculations from ICPSR behavioral observation data file, part of McPherson, M., Silloway, G., and Frey, D. [2006]. Crime, Fear, and Control in Neighborhood Commercial Centers: Minneapolis and St. Paul, 1970–1982. Inter-university Consortium for Political and Social Research [distributor].)*

Look at the level-2 relationship shown in Figure 8.4.2, in the scatterplot predicting the proportion of business-using pedestrians, with the proportion of young pedestrians as the predictor. As the proportion of young pedestrians increases, up to a value of about .27, so too does the proportion of business-using pedestrians. Nevertheless, at higher values of the proportion of young pedestrians, the proportion of business-using pedestrians starts to decline. In short, there is a slightly positive linear relationship between a younger age mix and a higher proportion of business-using pedestrians (line = long dash and dot) but a negative curvilinear relationship (dashed line) between the younger age mix and the business-using mix. The figure also shows a locally weighted robust regression line (Cleveland 1979) that follows the local data pattern. This roughly matches the overall curvilinear relationship except at the very lowest values on the horizontal axis.

Results appear in the Supplement 8.4 material, results section. Suppose you did just a plain monolevel logit model (starting line # 1254). The odds ratio for kidrteen (.54) means that the odds of a pedestrian being [primarily business-using vs. not primarily business-using] were about [1 − .54] 46 percent lower if the pedestrian was a child or a teen as compared to an adult or a senior ($p < .001$).

With a mixed version of this model, allowing each SCC to have its own outcome proportion (starting line # 1278), the negative impact of being a young pedestrian became a bit stronger (odds ratio = .48). The relationship strengthened slightly because each SCC was allowed its own average outcome score. The relationship, however, was

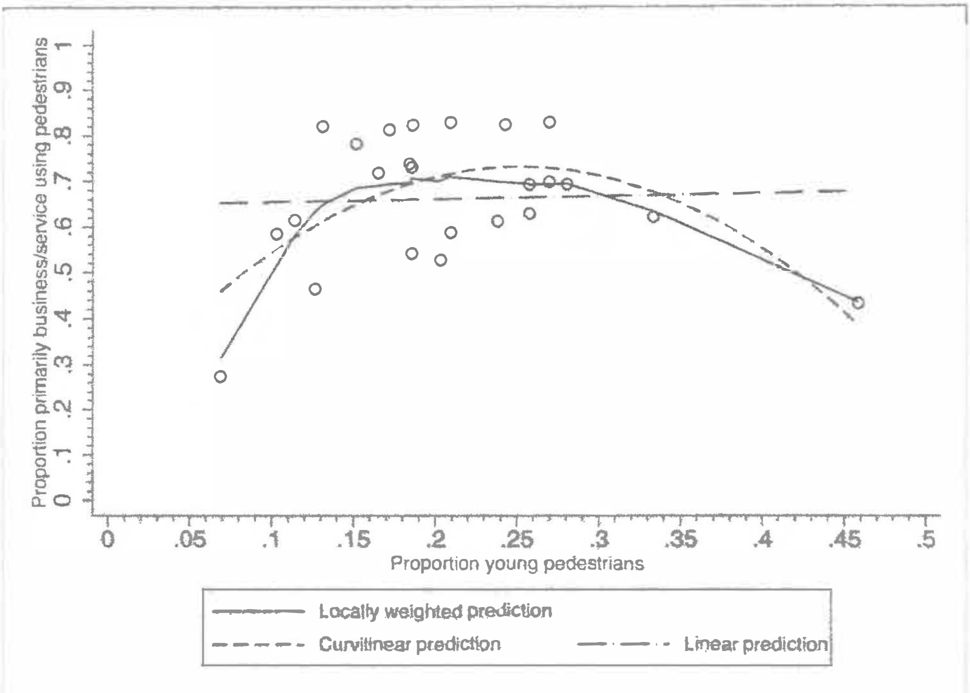

Figure 8.4.2 Scatterplot of relationship between proportion of young pedestrians (horizontal axis) and proportion of business-using pedestrians (vertical axis). Three statistical relationships shown: linear, curvilinear, and locally weighted robust regression line. *(Data source: ICPSR behavioral observation data file, part of McPherson, M., Silloway, G., and Frey, D. [2006].* Crime, Fear, and Control in Neighborhood Commercial Centers: Minneapolis and St. Paul, 1970–1982. *Inter-university Consortium for Political and Social Research [distributor].)*

still negative. But, there is a catch. The impact still bundles up, in ways you cannot separate, the between-pedestrian-within-SCC link to the outcome and the between-SCC link to the outcome.

How to separate these out? Split the predictor into two separate ones, a between-pedestrian-within-SCC or level-1 version and a between-SCC or level-2 version. The level-2 version is just the proportion of young pedestrians in each SCC (x_kidrteen). The level-1 version is created by group mean centering. Subtract the SCC-mean on proportion young, the values shown along the x-axis in Figure 8.4.2, from each pedestrian's individual score of 1 (young) or 0 (not young). A young pedestrian (scoring 1) at 38th and 4th in Minneapolis, after subtracting the SCC mean on this variable, .458, now scores .54. An older pedestrian at this center (scoring 0) now scores −.458 after this centering operation. Location matters because the variable is now social psychological, contrasting the individual with his or her group average. Take the 468 pedestrians observed at 15th and Nicollet in Minneapolis, where only 6.8 percent of observed pedestrians were young. On the new "level-1 group-mean-centered age" variable, a young pedestrian scores (1 − .068 =) .932, thus "sticking out" much more, when these scores are pooled across SCCs, than a young pedestrian at 38th and 4th. The new group-mean-centered variable captures how much younger a pedestrian was *compared to* his or her respective group in his or her SCC. The new variable averages zero across all SCCs, thus capturing only level-1 variation.

The level-1, social-psychological, group-mean-centered predictor (wi_kidrteen) and the level-2 ecological (x_kidrteen) variable are completely independent of each other ($r = 0$). You can now pull apart the two different relationships at the two different levels if you enter both predictors in a model.

Additional modifications are performed on the SCC-level age variable. The SCC average is centered by the overall average, so its mean is zero (c_x_kidrteen). Then this variable is squared to create a predictor to capture the curvilinear impact (sq_x_kidrteen). Given what you saw on the level-2 scatterplot, you would expect a positive linear impact of SCC proportion young pedestrians and a negative curvilinear impact of the squared variable, reflecting the downward bend at the right.

And, you get this (results starting line # 1379). The level-1, within-SCC impact remains relatively unchanged (OR = .47; $p < .001$), but the *interpretation* has shifted. Younger pedestrians whose youth contrasted more strongly with the respective age mix in their particular SCCs were *less* likely—a negative relationship—to be business-using pedestrians than the older pedestrians who contrasted more strongly with the age mix in their respective SCCs. The level-2, between-SCC linear impact of proportion young proved positive and significant ($z = 2.32$; $p < .05$), and the level-2, between SCC curvilinear impact of proportion young proved negative and significant ($z = -3.66$; $p < .001$). Translating level-2 odds ratios into practical terms requires discussing the impacts of standard deviation changes in the predictors (Long and Freese 2006).

What have you learned? You started with a monolevel model showing a sizable negative impact: younger pedestrians were less likely to be classified as business-using pedestrians. This aligned with the findings from situational action theory (Wikstrom et al. 2012) showing youths gathering for unsupervised and sometimes crime-promoting activities in SCCs. But, after multileveling the predictor, the mixed model showed three different and significant impacts of pedestrian age: one social psychological, and two ecological. (1) Within SCCs, the younger pedestrians whose youth stood out more given the age mix in the center, were *less* likely to use businesses or services (social psychological dynamic in a negative direction). (2) Across SCCs, the greater the fraction of younger pedestrians observed, the greater the fraction of business-using pedestrians observed (positive linear ecological dynamic). (3) But, past a certain point, this ecological relationship reversed (negative curvilinear ecological dynamic).

Advantages and Disadvantages

What Are the Main Advantages?

Policy and Practice

Mixed models help identify at what level the relevant dynamics are taking place. In this instance, differences within SCCs proved important, as did differences across SCCs. From an intervention perspective, a practitioner or a policy maker needs to consider both arenas.

Theoretical

Identifying levels at which processes are operating is crucial theoretically as well (Taylor 2015). Results showed dynamics working *in different directions* at different levels. Social psychological and ecological dynamics both merit theoretical attention.

What Are the Main Disadvantages?

Practical

You will find, once you start deploying these models, that things get complicated quickly. It is easy to get lost in the sauce and/or misled. It is strongly recommended that you model as much as you can using monolevel models as a first step. That way you at least have an estimate of what to expect before you start with the mixed models. Doing this will give you something solid to turn back to as needed. Advanced texts provide worked examples and guidelines for model building sequences. Work along with those.

Theoretical

To riff on the late Shel Silverstein's poem "Where the Sidewalk Ends," you can run out of theoretical sidewalk quickly. Take the modest example here. Situational action theory (Wikstrom et al. 2012) undergirds the negative social psychological impact observed. But the two ecological impacts lack a theoretical frame. We have run out of theoretical sidewalk. Results like those observed here, if they replicate, should spur scholars to elaborate the available multilevel *theoretical* models or develop new ones.

Next Steps

Read. Start out with the simple stuff (Bickel 2006; Finch et al. 2019; Garson 2019; Luke 2004; Robson and Pevalin 2016). If you are ready for the harder stuff, many excellent advanced texts are available (Gelman and Hill 2007; Hox 2010; Rabe-Hesketh and Skrondal 2012a, 2012b; Raudenbush and Bryk 2002; Snijders and Bosker 2012). Once you have grasped the fundamentals presented here, you are ready to take on these weightier tomes. Rabe-Hesketh and Skrondal (2012a, 2012b) may prove especially helpful for readers relying extensively on Stata.

NOTES

1. Bayesian statistics is an entire universe unto itself. The term "Bayesian" comes from Thomas Bayes (1702–1761). He "was a reputable mathematician and Presbyterian minister in England" who invented Bayes' Theorem also known as Bayes' Rule (Kruschke 2011, p. 62). Bayesian inference "gets us from prior to posterior beliefs" where prior beliefs are what we believe before we examine the data at hand, and posterior beliefs are what we believe after examining those data (Kruschke 2011, p. 12). How do our views about something change when we have additional data relevant to our beliefs? For example, we might believe a single die will end a roll with a "one" facing up about 16–17 percent or one-sixth of the time. This would be our prior belief. But, if we roll it 100 times, and it comes up "one" 50 percent of the time, our posterior belief about this die, and its fairness, would be different. Bayes and Laplace in France "receive independent credit as the first to *invert* the probability statement and obtain probability statements about θ [an unobserved parameter], *given* observed y" (Gelman et al. 2003, p. 34). Instead of saying "how likely are these data?" a Bayesian asks: "*Given* these data, how likely is my model feature?" There are full Bayesian statistics (Congdon 2006; Gelman et al. 2003) and empirical Bayesian statistics. The latter is of concern here.

2. The term "overall mean" is used loosely here: This analysis is considering a "precision weighted average" as the appropriate value for the overall mean, G_{00}, not the arithmetic mean (Raudenbush and Bryk 2002, p. 40). Precision weighting takes into account that "the residual standard deviations of some cases [group means here] are larger than for others, and estimates will be more precise if cases with higher residual standard deviation get lower weight" (Snijders and Bosker 2012, p. 220).

3. The graph suggests that the SCC with the highest score on proportion of young pedestrians might be an outlier, having an outsize influence on the linear and curvilinear relationships. One could investigate further by Winsorizing (Tukey and McLaughlin 1963) the highest value, and rerunning the model.

REFERENCES

Bauer, D. J. (2003). Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics, 28*(2), 135–167.

Baumer, E. P., and Arnio, A. N. (2012). Multi-level modeling and criminological inquiry. In D. Gadd, S. Karstedt, and S. F. Messner (Eds.), *The SAGE Handbook of Criminological Research Methods* (pp. 97–110). Thousand Oaks, CA: Sage.

Bickel, R. (2006). *Multilevel analysis for applied research: It's just regression*. New York: Guilford.

Blasko, B. L., Roman, C. G., and Taylor, R. B. (2015). Local gangs and residents' perceptions of unsupervised teen groups: Implications for the incivilities thesis and neighborhood effects. *Journal of Criminal Justice, 43*(1), 20–28. Available at https://doi.org/10.1016/j.jcrimjus.2014.11.002.

Browne, W. J., Goldstein, H., and Rabash, J. (2001). Multiple membership multiple classification (MMMC) models. *Statistical Modelling, 1*, 103–124.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association, 74*, 829–836.

Congdon, P. (2006). *Bayesian statistical modeling* (2nd ed.). West Sussex, U.K.: Wiley.

Covington, J., and Taylor, R. B. (1991). Fear of crime in urban residential neighborhoods: Implications of between and within-neighborhood sources for current models. *Sociological Quarterly, 32*, 231–249.

Finch, W. H., Bolin, J. E., and Kelley, K. (2019). *Multilevel modeling using R* (2nd ed.). New York: CRC.

Garcia, R. M., Taylor, R. B., and Lawton, B. A. (2007). Impacts of violent crime and neighborhood structure on trusting your neighbors. *Justice Quarterly, 24*(4), 679–704. Available at https://doi.org/10.1080/07418820701717169.

Garson, G. D. (2019). *Multilevel modeling: Applications in STATA, IBM SPSS, SAS, R, and HLM*. Thousand Oaks, CA: Sage.

Gelman, A., Carlin, J. B., and Stern, H. S. (2003). *Bayesian data analysis* (2nd ed.). Boca Raton, FL: Chapman and Hall/CRC.

Gelman, A., and Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Haberman, C. P., Groff, E. R., and Taylor, R. B. (2013). The variable impacts of public housing community proximity on nearby street robberies. *Journal of Research in Crime and Delinquency, 50*(2), 163–188. Available at https://doi.org/10.1177/0022427811426335.

Hox, J. J. (2010). *Multilevel analysis* (2nd ed.). East Sussex, U.K.: Routledge.

Kruschke, J. K. (2011). *Doing Bayesian data analysis*. Burlington, MA: Elsevier.

Liska, A. E. (1990). The significance of aggregate dependent variables and contextual independent variables for linking macro and micro theories. *Social Psychology Quarterly, 53*(4), 292–301.

Lockwood, B. (2011). The influence of travel distance on treatment noncompletion for juvenile offenders. *Journal of Research in Crime and Delinquency*. Available at https://doi.org/10.1177/0022427811414198.

Long, J. S., and Freese, J. (2006). *Regression models for categorical dependent variables using Stata* (2nd ed.). College Station, TX: Stata Press.

Luke, D. A. (2004). *Multilevel modeling*. Thousand Oaks, CA: Sage.

Maxfield, M. G. (1984). The limits of vulnerability in explaining fear of crime: A comparative neighborhood analysis. *Journal of Research in Crime and Delinquency, 21*(3), 233–250.

McPherson, M., Silloway, G., and Frey, D. (2006). *Crime, fear, and control in neighborhood commercial centers: Minneapolis and St. Paul, 1970–1982*. Minneapolis: Minneapolis Crime Prevention Center.

Perkins, D., and Taylor, R. B. (1996). Ecological assessments of disorder: Their relationship to fear of crime and theoretical implications. *American Journal of Community Psychology, 24*, 63–107.

Rabe-Hesketh, S., and Skrondal, A. (2012a). *Multilevel and longitudinal modeling using Stata, Vol. 1: Continuous Responses* (3rd ed.). College Station, TX: Stata Press.

———. (2012b). *Multilevel and longitudinal modeling using Stata, Vol. 2: Categorical Responses, Counts and Survival* (3rd ed.). College Station, TX: Stata Press.

Raudenbush, S. W., and Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Robinson, J., Lawton, B., Taylor, R. B., and Perkins, D. D. (2003). Longitudinal impacts of incivilities: A multilevel analysis of reactions to crime and block satisfaction. *Journal of Quantitative Criminology, 19*, 237–274.

Robson, K., and Pevalin, D. (2016). *Multilevel modeling in plain language*. Los Angeles, CA: Sage.

Sampson, R. J. (2012). *Great American city: Chicago and the enduring neighborhood effect*. Chicago: University of Chicago Press.

Sampson, R. J., Raudenbush, S. W., and Earls, F. (1997). Neighborhoods and violent crime: A multi-level study of collective efficacy. *Science, 277*, 918–924.

Snijders, T.A.B., and Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Los Angeles, CA: Sage.

Sorg, E. T., and Taylor, R. B. (2011). Community-level impacts of temperature on urban street robbery. *Journal of Criminal Justice, 39*(6), 463–470. Available at https://doi.org/10.1016/j.jcrimjus.2011.08.004.

Taylor, R. B. (2001). *Breaking away from broken windows: Evidence from Baltimore neighborhoods and the nationwide fight against crime, grime, fear and decline*. New York: Westview.

———. (2015). *Community criminology: Fundamentals of spatial and temporal scaling, ecological indicators, and selectivity bias*. New York: New York University Press.

Taylor, R. B., Harris, P. W., Jones, P. R., Garcia, R. M., and McCord, E. S. (2011). Ecological origins of shared perceptions of troublesome teen groups: Implications for the basic systemic model of crime, the incivilities thesis, and political economy. *Journal of Research in Crime and Delinquency, 48*(2), 298–324. Available at https://doi.org/10.1177/0022427810391537.

Taylor, R. B., Kelly, C. E., and Salvatore, C. (2010). Where concerned citizens perceive police as more responsive to troublesome teen groups: Theoretical implications for political economy, incivilities and policing. *Policing and Society, 20*(2), 143–171.

Taylor, R. B., and Lawton, B. A. (2012). An integrated contextual model of confidence in local police. *Police Quarterly, 15*(4), 414–445. Available at https://doi.org/10.1177/1098611112453718.

Tukey, J. W., and McLaughlin, D. H. (1963). Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/Winsorization 1. *Sankhyā: Indian Journal of Statistics, Ser. A (1961–2002), 25*(3), 331–352. Available at https://doi.org/10.2307/25049278.

Wikstrom, P.-O. H., Oberwittler, D., Treiber, K., and Hardie, B. (2012). *Breaking rules: The social and situational dynamics of young people's urban crime*. Oxford: Oxford University Press.