Chapter twenty

Multilevel Regression Models

How do we analyze clustered or multilevel data?

Variance components model

What are Fixed and Random Effects?
How Do We Interpret the Results from a Variance Components Model?
What is the Intraclass Correlation and How is it Interpreted?

Random intercept model

How is the Random Intercept Model Different from the Variance Components Model?
How is Explained Variance Computed and Interpreted?
What are Between and Within Effects?
How Do We Test for Differences in Between and Within Effects?

Random coefficient model

How is the Random Coefficient Model Different from the Random Intercept Model?
How Do We Test for Differences Between the Random Coefficient Model and the Random Intercept Model?

Adding cluster (level-2) characteristics

How are Cluster (Level-2) Characteristics Added to the Random Coefficient Model?
How Do We Interpret the Effects of Cluster (Level-2) Characteristics?

D. Weisburd and C. Britt, *Statistics in Criminal Justice*, DOI 10.1007/978-1-4614-9170-5_20,

Throughout the discussion of regression models—both OLS and logistic—we have assumed a single sample of cases that represents a single sample frame. What happens if there are clusters of cases in our data? For example, it is common in community-based surveys to first select a set of neighborhoods from a larger population of neighborhoods. Then, within these neighborhoods, a sample of individuals is selected to respond to the survey questions. A second example may involve researchers collecting data from a survey of youth by first sampling schools and then administering the survey to students within the selected schools, which may be further complicated by selecting a limited number of classrooms within each school. A third example might involve an experiment with multiple treatment and control sites, but the statistical analysis only distinguishes treatment from control group, ignoring information about site. This kind of design, where there are different levels of sampling, has the potential to allow the researcher to look at the effects of the larger sampling unit (e.g., classroom, neighborhood, school, or research site) on individual responses or outcomes.

Clustered data is also conceptualized as **multilevel data** (and hence, the name of the statistical models we discuss in this chapter). In thinking about clustered data as multilevel data, we would define the cluster—neighborhood, school, treatment site above—as the level 2 data. The unique observations—typically, individual cases—would be defined as the level 1 data. We are also not limited to thinking of our data as having only two levels and could conceivably work with data that have three or more levels. An example of three-level data might involve a study that begins with a sample of schools (level 3), followed by a sample of classrooms within each of those schools (level 2), and then the individual students within each of the selected classrooms (level 1). Put in terms of clustered data, we have students clustered within classrooms that are clustered within schools. The more general point to describing clustered data as multilevel data is that the lowest level of data will represent the total number of observations in our data—whatever these observations happen to represent. Each additional level of clustering then reflects a higher level of data.

Why does the clustering of data matter? There are both statistical and theoretical reasons for why we may want to pay attention to clustered data. Statistically, observations within a cluster will tend to be more similar to each other than to observations from different clusters. For example, survey respondents within a neighborhood will tend to be more alike on key individual characteristics when compared to survey respondents from another neighborhood, regardless of whether that other neighborhood is across town or across the nation. The increased similarity of cases within a cluster has consequences for our statistical tests, making it more likely that we will find statistically significant results, since cases within a cluster will tend to exhibit a similar pattern of association and consequently smaller standard errors.[1]

Theoretically, we may also have an interest in the multilevel structure of the data that points to important effects of the cluster on relationships observed at the individual level. For example, how might characteristics of a neighborhood—such as poverty rate or unemployment rate—affect the relationship we might observe between a respondent's gender and fear of crime? If we find that female respondents express higher levels of fear of crime, then we could ask the question about whether this statistical relationship is the same across neighborhoods. Does the effect of gender on fear of crime change across neighborhood? If the effect is essentially the same across neighborhood, it tells us that neighborhood may be unimportant for understanding fear of crime. In contrast, if we find the effect of gender does vary across neighborhood, we may then want to investigate why the effect varies. Is it due to other characteristics of the neighborhood, such as poverty, unemployment, vacant houses, and the like? Multilevel data are structured in such a way that the clustering of cases presents both a challenge and an opportunity to test for the effects of different independent variables measured for different units of analysis.

In this chapter, we provide a brief introduction to what are known as multilevel models[2] that account for the clustering of cases—the multilevel structure of the data—and can tell us interesting things about the nature of the statistical relationships we are studying. We take as a given that there is something informative or interesting about the multilevel structure of the data—that the clustering of observations is not simply a statistical nuisance to be corrected. In the discussion that follows, we restrict our attention to the analysis of dependent variables measured at the interval or ratio level of measurement. We also limit our discussion to two-level models: we have individual-level data (level 1) nested within one set of clusters (level 2). There is an extensive and growing literature on increasingly more sophisticated multilevel models that account for dependent variables measured at the nominal and the ordinal levels of measurement as well as multilevel data with three or more levels.[3] These models are too complex to examine in this brief introductory treatment of multilevel models.

---

[1] If our concern is primarily in statistically accounting for clustered observations, we can use what are referred to as robust standard errors and is available as an option in most statistical packages. We do not discuss these standard errors in this chapter, but encourage curious readers to consult Angrist, J.D., & Pischke, J. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

[2] These models are also known as mixed models, random effects models, and hierarchical linear models. Since these phrases all take on different meanings across the social and behavioral sciences, we use multilevel models, since that phrase seems to hold less potential for confusion across disciplinary boundaries.

[3] See, for example, Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods*, 2nd edn. Thousand Oaks, CA: Sage.

Variance Components Model

A Regression Approach to Analysis of Variance

We begin our discussion of multilevel models by starting with the simplest case, that of assessing how much cluster (i.e., group) means vary from each other. We find that the most straightforward building block for accomplishing this is the simple one-way analysis of variance model. Recall from our discussion of analysis of variance in Chap. 12 that our presentation emphasized how the decomposition of the dependent variable's total variance into two parts—between-group and within-group—could be used to test whether group had any ability to "explain" (statistically) the total variation in the dependent variable. Fundamentally, we were trying to assess whether the group means were significantly different from each other.

As a linear statistical model similar to OLS regression, we note that a one-way analysis of variance can be written in equation form analogous to an OLS regression equation:

$$Y_{ij} = b_0 + b_j X_{ij} + \epsilon_{ij},$$

where Y_{ij} is the value of the dependent variable for individual i in group j, X_{ij} represents the group (j) that an individual (i) belongs to, b_0 is a mean of the dependent variable, b_j a measure of the distance between each group and b_0, and ϵ_{ij} is the individual residual. The specific meaning of b_0 and each b_j depends on how the X_{ij} have been coded, which we explain below. As in the OLS regression model, the error term (ϵ_{ij}) is assumed to have a normal distribution with a mean of 0 and variance of σ_e^2.

There are two primary ways that the X_{ij} are typically coded to estimate this model—regression coding and contrast coding. Regression coding refers to creating a series of dummy variables coded 0–1 for all groups except one, which is used as the reference category. This procedure is identical to that discussed in Chap. 16 on the use of multiple dummy variables to represent a multi-category nominal independent variable in OLS regression analysis. If regression coding is used, we could rewrite the above equation as

$$Y_{ij} = b_0 + b_1 D_{1i} + b_2 D_{2i} + \ldots + b_{(j-1)} D_{(j-1)i} + \epsilon_{ij},$$

where each D_j represents a dummy variable indicator for up to $j - 1$ of the j groups (since one group does not have an indicator variable and functions as the reference category). In this model, b_0 represents the mean for the omitted group and each b_j measures the difference between the mean for group j and the mean for the omitted group.

Contrast coding works in a similar way to dummy variable coding, with the difference being the reference category in the regression coding scheme (i.e., the category with a 0 for each dummy variable) is coded as a −1 in contrast coding. Contrast coding ensures that the sum of all the estimated effects (i.e., the b_j) is 0, meaning that we can always determine the value for the reference category. We could rewrite the one-way ANOVA equation using contrast coding as

$$Y_{ij} = b_0 + b_1 C_{1i} + b_2 C_{2i} + \ldots + b_{(j-1)} C_{(j-1)i} + \epsilon_{ij},$$

where each C_j represents the contrast coded indicator for up to $j - 1$ of the j groups. In this model, b_0 represents the overall sample mean and the b_j measure of the distance between each group mean and the overall sample mean. The distance between the group mean for the category coded as -1 for each C_j is simply the negative of the sum of the other effects.

For example, suppose we have an experiment with three conditions: Treatment 1, Treatment 2, and Control Group. If we designate the Control Group as the reference category, the contrast coding scheme would look like the following:

	C_1	C_2
Treatment 1	1	0
Treatment 2	0	1
Control Group	–1	–1

The value for the Control Group would be $-(C_1 + C_2)$. Suppose that we estimated this model and found the following:

$b_0 = 2.3$, $C_1 = 0.3$, $C_2 = 0.6$:
Mean for Treatment 1: $2.3 + 0.3 = 2.6$
Mean for Treatment 2: $2.3 + 0.6 = 2.9$
Mean for Control Group: $2.3 - (0.3 + 0.6) = 1.4$

A Substantive Example: Bail Decision-Making Study

One of the most important decisions in the criminal process is the bail and release decision made by the judge shortly after the arrest of most individuals. Several decades of research have shown that defendants who have been released during the pretrial period—the time between arrest and disposition (conclusion) of a case—will tend to receive more lenient punishments if convicted. Those defendants who remain in jail during the pretrial period, due to the judge denying release altogether or to the judge requiring a bail amount the defendant could not pay, will typically be more likely to go to prison and to receive slightly longer sentences if sentenced to prison. Importantly, these effects hold, even after taking into account other characteristics of the defendant, such as prior record and severity of the offense.

As part of a larger project exploring judicial decision-making in the bail and release decision, John Goldkamp and Michael Gottfredson conducted two studies in Philadelphia—the first a pilot study to examine the factors that influenced the level of bail judges required and the second a test of whether the use of what were called bail guidelines made the decision-making more consistent and equitable across defendants.[4] In this chapter, we focus our attention on data from the first study. Goldkamp and Gottfredson selected a random sample of 20 judges to participate in the study and then selected a random sample of 240 cases per judge that required a bail and/or release decision. This resulted in a total sample of 4,800

[4] Goldkamp, J.S., & Gottfredson, M.R. (1985). *Policy guidelines for bail: An experiment in court reform*. Phildelphia, PA: Temple University Press.

cases clustered evenly across the 20 judges. Put in the terminology of levels of data, the 4,800 cases represent our level 1 data, while the 20 judges represent our level 2 data.

We can consider each judge as a separate experimental condition—cases were randomly assigned to each judge, ensuring broad similarity of the cases and thereby creating the opportunity to assess how similarly or differently judges would process these cases. Our attention in the example that follows is the bail decision for each case that was indicated by the dollar amount the judge set for the person's release.[5] Our dependent variable is the common logarithm of bail amount. Of the original 4,800 cases, bail amounts were required of 2,314 defendants, which comprise the sample for the following analyses.

Table 20.1 presents the means and standard deviations for bail amount (in dollars) and logged bail amount for each of the 20 judges. The average bail amount required by each judge varies considerably. For example, the average bail amount required by Judge 9 was \$1,652, while for Judge 6, it was \$17,487. Note that the values for logged bail are much smaller and have a more limited range. This is due to the fact that the logarithm used here—base 10—reflects the exponent for the number of times 10 would be multiplied by itself to reproduce the bail amount (e.g., log(100) = 2, log(1000) = 3, and so on). Consequently, a difference of 1.0 on the logarithmic scale used here is equivalent to a ten-fold increase in bail amount.

To help establish a baseline for the multilevel models discussed in the remainder of this chapter, it will be useful to present the results from a one-way ANOVA, where we treat each judge as a type of experimental condition and test for differences in mean bail amounts. Table 20.2 presents the results for logged bail amount (since this will be the outcome measure we rely on in subsequent analyses in this chapter). Note that for the coefficients reported in Table 20.2, we have used dummy variable coding, using Judge 1 as the reference category. Consequently, the model intercept represents the mean for Judge 1 and each of the reported coefficients represents the difference between that judge's mean and that for Judge 1. This can easily be confirmed by noting the coefficient for Judge 2: 0.24. The mean for Judge 2 is thus $3.04 + 0.24 = 3.28$, which is identical to the value reported in Table 20.1. Similarly, for Judge 10, the coefficient is −0.08, so the mean for Judge 10 is $3.04 - 0.08 = 2.96$.[6] The test of the null hypothesis of equality of means across judge gives us an F-test value of $F = 14.48$ with $df_1 = 19$, $df_2 = 2294$, and $p < 0.001$. We then conclude that the mean logged bail amount across this sample of 20 judges is significantly different.

Fixed and Random Effects

In the analysis of variance model as we have presented it above, the b_j are referred to as **fixed effects**, meaning that they represent a constant effect of the group for all of the cases within that group. In experimental research, this implies that the treatment received by each individual assigned to a particular condition is assumed

[5] There was a 10% rule in effect in Philadelphia at the time of the study, meaning that defendants would only need to post 10% of the dollar amount requested by the judge in order to ensure their freedom during the pretrial period.

[6] Due to rounding, some of the judge means estimated with the coefficients in Table 20.2 will differ at the second decimal when compared to the means reported in Table 20.1.

Table 20.1 Means and Standard Deviations of Bail Amounts by Judge in Philadelphia

	BAIL(DOLLARS)		BAIL(LOGGED)			BAIL(DOLLARS)		BAIL(LOGGED)	
JUDGE	MEAN	SD	MEAN	SD	JUDGE	MEAN	SD	MEAN	SD
1	2076.30	4513.85	3.04	0.45	11	4576.19	7170.58	3.38	0.47
2	4784.88	8522.53	3.28	0.56	12	7299.55	16270.59	3.34	0.63
3	1901.68	3414.13	3.05	0.41	13	3385.96	6310.05	3.21	0.50
4	1830.43	2808.85	2.97	0.46	14	7945.31	24293.52	3.32	0.65
5	2204.58	3890.20	2.98	0.50	15	4944.00	11726.94	3.28	0.51
6	17486.78	60861.05	3.51	0.73	16	8747.40	19082.44	3.57	0.50
7	1842.76	3320.16	3.00	0.42	17	2184.62	3182.93	3.11	0.41
8	2117.76	3583.45	3.01	0.47	18	4158.82	7765.01	3.24	0.51
9	1652.86	2099.00	2.95	0.46	19	1956.30	3053.54	3.07	0.38
10	3627.04	10960.82	2.96	0.57	20	6246.90	23285.99	3.23	0.58

Table 20.2 Analysis of Variance of Logged Bail Amount

JUDGE	COEFFICIENT (b_j)	STD. ERROR
2	0.24	0.07
3	0.01	0.06
4	–0.08	0.07
5	–0.06	0.06
6	0.47	0.06
7	–0.05	0.06
8	–0.04	0.07
9	–0.09	0.06
10	–0.08	0.07
11	0.33	0.07
12	0.30	0.07
13	0.16	0.07
14	0.28	0.06
15	0.24	0.06
16	0.53	0.07
17	0.07	0.07
18	0.19	0.07
19	0.03	0.06
20	0.18	0.06
Intercept	3.04	0.04

to be the same. However, this assumption ignores the fact that there are often differences in the kind of treatment each case within a particular condition may receive—known as treatment effect heterogeneity. Although the lead researcher may have designed a protocol that minimizes variations in the treatments received by participants, the requirement for many treatments to be administered by another human being introduces the possibility of differences in multiple administrations of the treatment. For example, an individual police officer may think that he/she is meeting the expectations of the researcher, but events and circumstances unique to that officer, that day, that site, may result in slight differences in how a treatment is administered.

If we assume that there are no systematic differences in how an individual administers a treatment within a group, then the analysis of variance model can be modified to incorporate **random effects**. These random effects allow for variation within a group or condition, which acknowledges that there will be differences in the treatments individuals in each group or condition receive. A parallel way of considering random effects is to think of the conditions of the group, whatever the group represents, as a sample of all possible conditions within the group.

The random effects model can be written as:

$$Y_{ij} = b_0 + \zeta_j + \epsilon_{ij}, \qquad \textbf{Equation 20.1}$$

where Y_{ij}, b_0, and ϵ_{ij} are as defined previously. The ζ_j (Greek letter zeta) are the random effects and represent the difference in mean for group j (as sampled) and the overall sample mean b_0. The ζ_j are assumed to have a normal distribution with a mean of 0 and variance of σ_z^2.

Note that we now have two measures of variance—σ_z^2 and σ_e^2—that reflect variation between groups (σ_z^2) and within groups (σ_e^2) and combined represents the total variation in the dependent variable:

$$\text{Var}(Y) = \sigma_z^2 + \sigma_e^2 \qquad \textbf{Equation 20.2}$$

These are what are known as the **variance components** that can be used to assess whether there is variation in the dependent variable across the group means.

How do we know when to choose a fixed or random effects model? Of primary consideration is whether the effect of the group is viewed as being consistent across all cases within the group or whether the effect of the group represents a sampling of all possible effects of the group. To the extent the effect of the group is viewed as consistent across cases, then a fixed effects model is the optimal choice and we would estimate a standard analysis of variance model. Alternatively, if the effect of the group is expected to vary among cases within that group, then a random effects model is the more appropriate choice.

From a practical standpoint, there are no firm rules about the sample sizes needed to estimate models with fixed and random effects. The total sample size (N) is used to estimate the fixed effects and much like estimating any linear regression model, relatively modest sample sizes (100–200 cases) are often adequate. That same guideline holds for multilevel models. Since the random effects are estimated at the level of the cluster, it is unclear just how many clusters are necessary to estimate a multilevel model, although 10–20 clusters provide a lower bound[7].

Intraclass Correlation and Explained Variance

Given the two measures of variance—σ_z^2 and σ_e^2—we can compute a measure of explained variance (ρ):

$$\rho = \frac{\sigma_z^2}{(\sigma_z^2 + \sigma_e^2)}, \qquad \textbf{Equation 20.3}$$

[7] Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and longitudinal modeling using stata, volume I: Continuous responses*, 3rd edn. College Station, TXL: Stata Press.

where ρ has values ranging from 0 to 1 and measures the proportion of total variation in the dependent variable that is due to the group. At $\rho = 0$, the group explains none of the variation in the dependent variable, while at $\rho = 1$, the group explains all of the variation in the dependent variable.

An alternative interpretation of ρ is as the **intraclass correlation**, which indicates the level of absolute agreement of values within each group. By absolute agreement, we're trying to assess the extent to which the values within a group are identical. Recall from Chap. 14 in our discussion of Pearson and Spearman correlation coefficients, we were assessing relative agreement in cases. For the Pearson correlation, it was the relative agreement in values of two variables, while for the Spearman correlation, it was the relative agreement in ranks of values of two variables.

The intraclass correlation provides a measure that can be viewed in two different ways. In part, it provides a measure of intergroup heterogeneity by measuring how much of the total variation is due to the group. At the same time, it provides a measure of within group homogeneity by measuring how similar the values are within each group.

Statistical Significance

A natural question to arise in the application of random effects models is whether the random effect—the estimate of variance σ_z^2—is statistically significant. Substantively, this is a question about whether allowing for random variation around the overall mean adds anything statistically to the model over and above a fixed effects model.

To test the statistical significance of σ_z^2, we rely on a likelihood-ratio test, similar to that used in previous chapters. To compute the LR test for the variance component, we need two values of the log-likelihood: (1) log-likelihood for the ANOVA model and (2) log-likelihood for the random effects model (REM). The LR test is computed as

$$\chi^2 = -2(LL(\mathbf{ANOVA}) - LL(\mathbf{REM})).$$ **Equation 20.4**

The likelihood-ratio test statistic has a χ^2 sampling distribution with 1 degree of freedom. We then divide the observed level of statistical significance for the computed χ^2 by 2, since it is a test of variances, which can only take on positive values and effectively truncates the sampling distribution to positive values.

Bail Decision-Making Study

We return to our example from the Bail Decision-Making Study and present the results for a variance components model in Table 20.3. The model intercept is estimated to be 3.17, which is also the sample mean for logged bail amount. The variance of the groups (σ_z^2) is estimated to be 0.031, indicating the degree to which the group (i.e., judge) means vary around the full sample mean. The unexplained error variance (σ_e^2) is quite a bit larger and is estimated to be 0.27.

To what extent does the judge making the decision about bail affect the required amount? The intraclass correlation provides an indicator of the influence

Table 20.3 Variance Components Results for Logged Bail Amount

VARIABLE	COEFFICIENT	se	*z*-SCORE
Fixed Effect:			
Intercept	3.17	0.04	77.73
Random Effects:			
Intercept (σ_z^2)	0.031		
Error (σ_e^2)	0.270		
ρ	0.104		

of the judge and is estimated to be 0.10 for logged bail. The intraclass correlation can also be obtained from the two variance components estimates:

$$\rho = \frac{\sigma_z^2}{(\sigma_z^2 + \sigma_e^2)} = \frac{0.03}{0.03 + 0.27} = 0.10.$$

The value of the intraclass correlation suggests that the decision-making judge only accounts for about 10% of the variation in the logged bail amounts.

In regard to statistical significance, we find that the log-likelihood for the one-way ANOVA model is −1871.73 and the log-likelihood for the random effects model is −1777.35.

$$\chi^2 = 2((-1871.73) - (-1777.35)) = 188.76$$

Based on 1 degree of freedom, we find the critical χ^2, assuming a p-value < 0.05, to be 3.841. Our computed χ^2 of 188.77 has a p-value much less than 0.05, even before dividing it by 2, meaning the variance components model represents a significant improvement over the standard one-way ANOVA model. Substantively, these results indicate that the decision-making judge is important to understanding bail amount requested.

Random Intercept Model

We can extend the basic variance components model to include independent variables to estimate what is known as a **random intercept model (RIM)**. Alternatively, we could start with an OLS regression model and allow the intercept to vary randomly across cluster. Either way, we estimate a model that takes on the form

$$Y_{ij} = b_0 + b_1 X_{1ij} + \zeta_j + \epsilon_{ij} \quad \textbf{Equation 20.5}$$

where Y and X_1 represent the dependent and independent variables, respectively, b_0 and b_1 represent the model intercept and the effect of X_1 on Y , ζ_j represents the random effect of the cluster on the model intercept, and ϵ_{ij} represents the random error term for each individual observation.

In the random intercept model, the regression coefficients are interpreted in the same way as discussed previously—a unit change in the independent variable is expected to result in a change in the dependent variable equal to b_1 and the intercept is the expected value of the dependent variable if X_1 has a value of 0.

Explained Variance

With a random intercept model, we have three variations on explained variance that help us to understand the patterns of association in our multilevel data. Recall that the total variance in the dependent variable is $\text{Var}(Y) = \sigma_z^2 + \sigma_e^2$. Following the estimation of a random intercept model, we can compute the explained variance (R^2) of the dependent variable with the following equation:

$$R^2 = \frac{(\sigma_{z0}^2 + \sigma_{e0}^2) - (\sigma_{z1}^2 + \sigma_{e1}^2)}{\sigma_{z0}^2 + \sigma_{e0}^2}, \qquad \textbf{Equation 20.6}$$

where σ_{z0}^2 and σ_{z1}^2 represent the variance of the random effects for the intercept in the variance components model (subscripted with a 0) and the random intercept model (subscripted with a 1), respectively. The error variance (unexplained variance) in the variance components model and the random intercept are indicated by σ_{e0}^2 and σ_{e1}^2, respectively. Consistent with previous interpretations of R^2, a value of 0 indicates none of the variance in the dependent variable was explained, while a value of 1 would indicate that all of the variance was explained.

We can further decompose the total explained variance into each of the two levels of data: (a) explained variance at the level of the cluster (level 2) and (b) unexplained variance at the level of the individual observations (level 1). The explained variance at level 2 is:

$$R_z^2 = \frac{\sigma_{z0}^2 - \sigma_{z1}^2}{\sigma_{z0}^2}. \qquad \textbf{Equation 20.7}$$

The explained variance at level 1 is:

$$R_e^2 = \frac{\sigma_{e0}^2 - \sigma_{e1}^2}{\sigma_{e0}^2}. \qquad \textbf{Equation 20.8}$$

What do these level-specific measures of explained variance tell us? The level 2 explained variance (R_z^2) informs us how much of the random variation in cluster means found in the variance components model is due to the individual level characteristics of the observations (i.e., the set of independent variables we included in the random intercept model). A value of 0 indicates that none of the cluster-level variation was explained by the characteristics of the individuals included in the data. Conversely, a value of 1 indicates that all of the cluster-level variation was due to the characteristics of the individuals included in the data. Generally, the level 2 explained variance will tell us how much of the observed cluster-level variance is due to the composition of the clusters.

The level 1 explained variance (R_e^2) tells us how much of the error (residual) variance—the "unexplained" variance—was reduced by adding in a set of independent variables thought to be related to the dependent variable. A value of 0 indicates that none of the error variance was explained by the inclusion of the independent variables, while a value of 1 would indicate that all of the error variance was explained by the independent variables. The explained variance at level 1 is directly analogous to our discussion of R^2 in linear regression models in Chaps. 15–17, where we gain a sense of how well our set of independent variables statistically explains the values on the dependent variable.

Statistical Significance

The results from a random intercept model will lead to testing the statistical significance of both the effects of the independent variables included in the model and the use of the random intercept model over an OLS linear regression model. In regard to testing for the statistical significance of the effects of the independent variables (e.g., b_1 in the equation above), we would use the following familiar equation:

$$z = \frac{b}{se(b)},$$ **Equation 20.9**

where $se(b)$ is the standard error of the coefficient and z the test statistic assumed to have a normal distribution. Similar to our discussion of testing the statistical significance of individual coefficients in logistic regression models in Chaps. 18 and 19, the coefficients in a random intercept model are also expected to be normally distributed. The maximum likelihood estimation procedure for the random intercept model estimates standard errors that are adjusted for the clustered nature of the data. Depending on the particular data we are working with, the standard errors for the coefficients in a random intercept model will always be at least as large, but more likely larger, than those estimated in an OLS linear regression model with the same variables, since the clustered nature of the data has been taken into account during the estimation process.

The test of the random intercept model against the OLS regression model involves the use of the same type of likelihood-ratio test that we used in the test of the variance components model against a one-way ANOVA model. The goal of this test is to assess whether allowing the model intercept to vary randomly across clusters improves the fit of the statistical model. The test is:

$$\chi^2 = -2(LL(\mathbf{OLS}) - LL(\mathbf{RIM})).$$ **Equation 20.10**

As before, the likelihood-ratio test statistic has a χ^2 sampling distribution with 1 degree of freedom for the one variance estimate of the random effects. We then divide the observed level of statistical significance by 2, since it is a test of variances, which can only take on positive values and effectively truncates the sampling distribution to positive values. Substantively, this test will indicate whether the addition of a random intercept to a linear regression model makes a statistically significant contribution.

Centering Independent Variables

There are many instances where the interpretation of the model intercept is important to understand the implications of the estimated results. When we are confronted with an independent variable that has no meaningful zero-point, the meaning of the intercept is difficult to explain. One of the straightforward ways of dealing with this issue is to center the independent variable. What do we mean by center a variable? In general, centering refers to subtracting a measure of central tendency (e.g., a mean) from each raw score.

The two types of centering of independent variable that are often most useful for estimating multilevel models are (1) grand or overall mean centering and (2) cluster or group-based centering. In grand-mean centering, we subtract the overall sample mean of the independent variable from each observation

$$X_{ij} - \bar{X}_{..}$$

Note that the two periods in the subscript of $\bar{X}$ indicate that the mean is the same for each individual observation *i* and group *j*.

In cluster-based centering, we subtract the relevant cluster (group) mean from each observation

$$X_{ij} - \bar{X}_{.j}$$

Note that there is a period in place of the *i* subscript to indicate the value of $\bar{X}_{.j}$ is the same for each observation *i* in a given group *j*.

How does the inclusion of a centered variable, instead of the original independent variable, affect the interpretation of the results? The regression coefficient for the centered value of the independent variable is interpreted in exactly the same way: a unit change in X is expected to result in a change of b units in the dependent variable. Keep in mind that all we have done by centering a variable is shifted its location; its scale is unchanged and so the slope is unchanged (this was discussed in Chap. 15 when we introduced the bivariate linear regression model).

The difference in interpretation that comes from using centered variables is in the model's intercept. In the grand-mean centering case, the model intercept now represents the expected mean for a case that has a value equal to the overall mean in the sample for X (which is equal to a value of zero on a grand-mean centered variable). For the cluster-based centering, the model intercept now represents a weighted average of the cluster means.

The other implication for interpreting the results is focused on the variance component σ_z^2. When grand-mean centering has been used, σ_z^2 represents variation in the group means around the overall mean, identical to the case where no centering has been used. With cluster-based centering, σ_z^2 is interpreted as the variation of group means around the weighted average of cluster means.

When should centering be used? In general, centering an independent variable measured at the interval or ratio level of measurement aids in the estimation of multilevel models, particularly some of the more complex models that involve estimating interaction effects across levels of data (which we discuss below). Centering should not change the substance of the statistical results, since all that centering accomplishes is a shifting of the independent variable so that a value of 0 represents either the overall mean for the sample or the cluster mean for each group.

Dummy variables may also be centered. In this case, the overall or group means simply represent the proportion of cases in the full sample or the cluster that has the characteristic measured by the dummy variable.

Table 20.4 Regression Results for Logged Bail Amount on Number of Prior Drug Offenses

			RIM	RIM
VARIABLE	OLS	RIM	GRAND MEAN CENTERING	CLUSTER CENTERING
Fixed Effects:				
Intercept	2.69	2.71	3.17	3.17
(se)	(0.019)	(0.040)	(0.037)	(0.041)
(*t* or *z*-score)	(141.22)	(67.11)	(85.18)	(77.62)
Number of Prior				
Drug Offenses	0.12	0.11	0.11	0.11
(se)	(0.004)	(0.004)	(0.004)	(0.004)
(*t* or *z*-score)	(29.87)	(29.87)	(29.87)	(29.80)
Random Effects:				
Intercept (σ_z^2)		0.026	0.026	0.031
Intercept (σ_e^2)		0.192	0.192	0.192

Bail Decision-Making Study

To illustrate the application of random intercept models and the use of centering, we return to the Bail Decision-Making Study. In the interest of keeping the statistical model simple, we include only one independent variable predicting bail amount. Clearly, there are numerous characteristics of defendants and their cases that affect the bail decision. Our goal here is to illustrate how one would go about estimating and interpreting a random intercept model with and without a centered independent variable. We will develop a more complex model in a later section.

In making assessments about bail, the judge is expected to consider both the chances of the defendant fleeing the community and the potential threat to public safety. One indicator of a defendant's overall risk is the number of prior drug offenses in the person's criminal history record. In general, the greater the evidence of prior drug offending, the higher the perceived risk of some kind of pretrial misconduct and consequently higher bail amounts being requested from defendants.

Table 20.4 presents results from four different models: OLS, random intercept model (RIM) with no centering, RIM with grand mean centering, and RIM with cluster mean centering. The OLS model ignores the clustering of cases by judge and simply reports the effect of number of prior drug offenses on the amount of bail requested. We see from the results in Column 1 that a one-unit increase in the number of prior drug offenses increases logged bail by 0.12 units. Since talking about changes in logged units likely makes little intuitive sense, a more meaningful interpretation of this coefficient is in terms of percentage change in the logged outcome variable. Our observed coefficient of 0.12 for number of prior drug offenses can alternatively be interpreted as expecting bail amount to increase by 12% for each additional prior drug offense.

The results for the RIM with no centering appear in Column 2 of Table 20.4. Note that the intercept remains nearly the same as in the OLS model (2.69 v. 2.72, respectively) as does the effect of number of prior drug offenses (0.12 v. 0.11, respectively). For the two RIMs with centering, the estimates for the intercept and the effect of number of prior drug offenses are the same through two significant digits (but not beyond). Note that the estimate of the intercept is the overall mean for logged bail amount reported earlier, which is expected under the use of grand mean centering. The use of cluster mean centering estimates a model intercept that is the weighted average of the group means. The reason the estimates for the intercept are so similar in the two models using different types of centering is an artifact of the study design that used a balanced approach to select the same number of cases for each judge. The small variation in the number of cases per judge used in these analyses accounts for the minor differences that do appear in the intercept and the effect of number of prior drug offenses.

In regard to the effect of number of prior drug offenses, please note that the effect is the same for all three RIMs, regardless of whether no centering (Column 2), grand mean centering (Column 3), or cluster mean centering (Column 4) was used. This is to be expected, as centering an independent variable only shifts the distribution of cases and does not alter anything else about the values, meaning that the coefficient representing the effect of the independent variable should stay the same.

As we did with the variance components model, we can test whether the RIM offers a statistically significant improvement over the OLS regression model. This again requires a chi-square test relying on the difference in the log-likelihood values for the OLS model and the RIM. The log-likelihood for the RIM model is −1400.06 and the log-likelihood for the linear regression model is −1513.37.

$$\chi^2 = -2((-1513.37) - (-1400.06)) = 226.62$$

Based on 1 degree of freedom, we find the critical χ^2, assuming a p-value < .05, to be 3.841. Our computed χ^2 of 226.62 has a p-value much smaller than 0.05, regardless of whether it is divided by 2. This result indicates that the RIM offers a substantial improvement in the statistical model over the traditional linear regression model.

The explained variance for the RIM is based on variance estimates presented in Tables 20.3 and 20.4. We calculate the overall R^2 to be:

$$R^2 = \frac{(0.031 + 0.270) - (0.026 + 0.192)}{0.031 + 0.270} = 0.276.$$

This shows that the inclusion of a single independent variable—number of prior drug offenses—resulted in a model that explained 27.6 % of the variation in logged bail.

When we decompose R^2 by level of data, we find the explained variance at level 2 (the judge) to be:

$$R_z^2 = \frac{0.031 - 0.026}{0.031} = 0.161.$$

Meaning that 16.1 % of the variation across judges is due to the composition of the cases in their courtroom. Put another way, about 16 % of the variation across judges is due to the number of drug offenses that defendants in their courtroom possess prior to the current arrest.

Although we have only added a single independent variable, we find the explained variance at level 1 (error variance) to be:

$$R_e^2 = \frac{0.270 - 0.192}{0.270} = 0.289.$$

What this means is a single independent variable—number of prior drug offenses—explains nearly 29 % of the variation in the error variance.

What we have not yet addressed is how to determine which type of centering to use, or whether to use any centering at all. We turn our attention to answering this question in the next section.

Between and Within Effects

In our discussion of centering, we noted that centering variables can assist in estimating multilevel models. What this means is that the algorithms used by various statistical packages to estimate multilevel models perform better when using centered independent variables. Although the explanation for how these algorithms work goes beyond the focus of our text, we note that statistical packages that estimate multilevel models often require multiple iterations to come to a solution—the estimates of the intercept and the other coefficients.

We are still left with the question, then, of which method of centering to use. How do we make this determination? One of the issues that naturally arises in the study of clustered or multilevel data is whether the effects of the independent variables are the same across group as they are within group. For example, in the analysis of judicial bail decision-making, we might wonder whether the effect of number of prior drug offenses across judge—what we will call a **between effect**—is the same as the effect of number of drug offenses processed by each judge—what we will call a **within effect**. Conceptually, what we are attempting to get at is whether a regression model for each judge (the within regression for each judge) is parallel to a single regression line based on the means for each judge (the between regression for all judges). To the extent the slopes (coefficients) are parallel, there are similar between and within effects, meaning that each judge uses information on number of prior drug offenses in approximately the same way. To the extent the slopes differ, there are different between and within effects, indicating that judges weight information about number of prior drug offenses differently.

Figures 20.1 and 20.2 present a way of thinking hypothetically about similar and different effects. In each figure, the solid line represents the overall regression slope for the effect of X_1 on Y. The dashed lines represent the regression lines within each of the five clusters plotted. In Fig. 20.1, the between and within effects are parallel to each other. The different placement of the dashed lines represents the random effect of each cluster (ζ_j), where two clusters have positive random effects (and appear above the solid line), and three clusters have negative random effects (and appear below the solid line). In Fig. 20.2, the between and within effects are different—one slope is positive, one slope is essentially flat, while the remaining three slopes are negative. Although these figures are informative in highlighting similarities and differences in the between and within effects, it is impractical to plot out regression lines for many groups, and will instead rely on a statistical test for differences in the between and within effects.

Fig. 20.1 *Parallel Between and Within Effects*

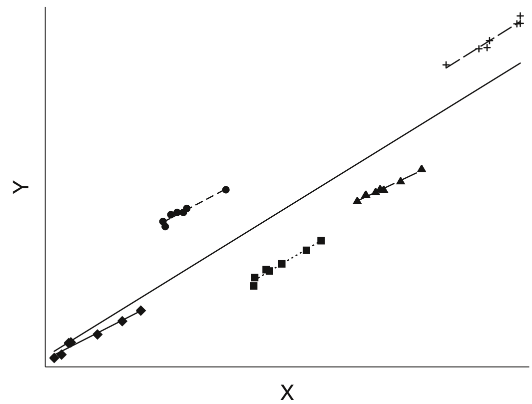
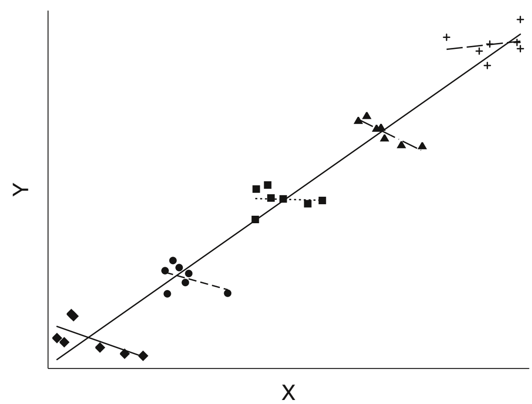


Fig. 20.2 *Different Between and Within Effects*



Testing for Between and Within Effects

The most direct way of testing for a difference in the between and the within effects is with the addition of the cluster mean for each independent variable already included in the random intercept model. There are two equivalent ways of testing for differences in the between and the within effects—both include an estimate for the cluster mean, but differ in whether the original raw score or the centered variable is included in the model. The following discussion illustrates the differences and equivalences between the two approaches.

First, we start with a simple random intercept model that has only a single independent variable:

$$Y_{ij} = b_0 + b_1 X_{1ij} + \zeta_j + \epsilon_{ij}.$$

We then add the cluster mean for X_1 and estimate the following equation:

$$Y_{ij} = b_0 + b_{1a} X_{1ij} + b_{1b} \bar{X}_{1.j} + \zeta_j + \epsilon_{ij}.$$ **Equation 20.11**

The coefficient b_{1b} estimates the magnitude of difference in the between and the within effects and captures any possible divergence in slope, such as that portrayed in Fig. 20.2. If b_{1b} is not significantly different from zero, then the between and the within effects are the same, and the slopes for each cluster parallel those of the overall effect. If b_{1b} is significantly different from zero, then the between and within effects diverge and the slopes are not parallel.

In the alternative, but fully equivalent, approach, we estimate a model that includes the cluster mean centered value for X_1 and the cluster mean for X_1 as two separate independent variables:

$$Y_{ij} = b_0 + b_{1a}(X_{1ij} - \bar{X}_{1.j}) + b_{1b}\bar{X}_{1.j} + \zeta_j + \epsilon_{ij}. \qquad \textbf{Equation 20.12}$$

In this model, the coefficient b_{1a} directly estimates the within effect and the coefficient b_{1b} directly estimates the between effect of X_1. To obtain the difference in between and within effects, we would simply subtract b_{1a} from b_{1b}.

Bail Decision-Making Study

In the Bail Decision-Making Study, one of the key areas of attention was a question about whether judges weighted information about defendants in similar or different ways. A direct test of this is provided by a test for similarity of between and within effects. If we continue the example started previously using number of drug offenses (*DRUGOFF*) as the independent variable, we estimate the following model:

$$\log(BAIL)_{ij} = b_0 + b_{1a}DRUGOFF_{1ij} + b_{1b}\overline{DRUGOFF}_{1.j} + \zeta_j + \epsilon_{ij}.$$

Table 20.5 presents the results from this model.

The key value presented in Table 20.5 is the estimate for the cluster means of number of drug offenses, which represents the difference of the within and the between effects. The within effects estimate has a value of $b = 0.11$. The estimate of the difference in the two effects has a value of $b = 0.35$, meaning the between effect is greater than the within effect, which confirms that the between and within effects of number of drug offenses are significantly different from each other. More importantly, in the context of understanding judicial decision-making, these results imply that the 20 judges in this study differentially weight information about number of prior drug offenses.

If we were interested in estimating both the within and between effects directly, we could use the second approach described above and estimate the following equation:

$$\log(BAIL)_{ij} = b_0 + b_{1a}(DRUGOFF_{1ij} - \overline{DRUGOFF}_{1.j}) + b_{1b}\overline{DRUGOFF}_{1.j} + \zeta_j + \epsilon_{ij}.$$

Table 20.5 Test of Between and Within Effects Similarity

VARIABLE	ESTIMATE	se	z-SCORE
Intercept	1.29	0.523	2.47
Number of Drug Offenses	0.11	0.004	29.80
Number of Drug Offenses (Cluster means)	0.35	0.128	2.72

Table 20.6 Test of Between and Within Effects Similarity

VARIABLE	ESTIMATE	se	z-SCORE
Intercept	1.29	0.523	2.47
Number of Drug Offenses (Cluster deviations)	0.11	0.004	29.80
Number of Drug Offenses (Cluster means)	0.46	0.128	3.61

Table 20.6 presents the results from this analysis.

As expected, given the previous set of results, the between effect of number of drug offenses is greater than the within effect of number of drug offenses, again confirming that these 20 judges differentially used information about drug offending when making bail decisions. There are two additional findings in Table 20.6 worth noting. First, the effect of cluster deviations is the same as the raw score estimates presented in Tables 20.4 and 20.5. This is to be expected, since this is just the effect for the cluster-mean centered number of drug offenses. Second, the difference in the between and the within effects is $0.46 - 0.11 = 0.35$, which is the estimate for the difference obtained directly using the first method (Table 20.5).

Random Coefficient Model

A straightforward extension of the random intercept model involves thinking about the effects of one or more of the independent variables in a multilevel model also varying across cluster. Put another way, we may have justification, based on prior research and theory, to expect the slope coefficients for a key variable to vary across cluster. For example, in a study of fear of crime across neighborhoods, we might expect the effect of gender to vary by neighborhood. Similarly, in our study of judicial decision-making, we might expect judges to weight information about cases and defendants differently, suggesting that we will find different slopes for key predictors of the outcome variable.

The development of the **random coefficient model (RCM)** begins with the random intercept model (here we have included only a single independent variable X_1):

$$Y_{ij} = b_0 + b_1 X_{1ij} + \zeta_{0j} + \epsilon_{ij}, \qquad \textbf{Equation 20.13}$$

where all terms are defined as above, except that there is now a 0 included in the subscript of the random effect ζ, to indicate the connection to the intercept (b_0). For a random coefficient model, we add a random effect for the slope coefficient in question. In our example, to estimate a model in which b_1 is allowed to vary across cluster, we would add an additional random effect ζ_{1j}:

$$Y_{ij} = b_0 + b_1 X_{1ij} + \zeta_{0j} + \zeta_{1j} X_{1ij} + \epsilon_{ij},$$

We can rewrite this equation to more directly link the random effects with the proper slope coefficient:

$$\begin{aligned} Y_{ij} &= (b_0 + \zeta_{0j}) + (b_1 X_{1ij} + \zeta_{1j} X_{1ij}) + \epsilon_{ij} \\ &= (b_0 + \zeta_{0j}) + (b_1 + \zeta_{1j}) X_{1ij} + \epsilon_{ij}. \end{aligned}$$

In this model, we estimate fixed effects for the model intercept (b_0) and the slope coefficient for X_1 (b_1) and simultaneously estimate random effects for both the intercept (ζ_{0j}) and the slope coefficient (ζ_{1j}).

Conceptually, what the random coefficient model does is analogous to estimating a regression model for each cluster and then examining whether the intercepts and slope coefficients vary in any meaningful way across the clusters. We now turn to a more formal examination of variance estimates from the random coefficient model.

Variance Estimates

Similar to the variance components and random intercept models, the level 1 error variance continues to be represented by σ_e^2. The variance of the random effects (the ζ) now takes on multiple values:

Variance of the intercept across cluster: σ_{z00}^2
Variance of the slope coefficient across cluster: σ_{z11}^2

We are now confronted with another choice in regard to the estimation of the variance components. In addition to the two variance estimates for the ζ, we may estimate the covariance of the random effects for the intercept and the slope, which we can label either σ_{z01}^2 or σ_{z10}^2. What does this estimate of the covariance of ζ_0 and ζ_1 assess? In general, it will indicate whether the magnitude of the random effect for the intercept covaries with the magnitude of the random effect for the slope coefficient. More directly, a positive value of σ_{z01}^2 indicates that clusters with larger intercepts will tend to have larger values for the slope coefficient. Conversely, a negative covariance would suggest that smaller values of the intercept are associated with larger values of the slope coefficient, and vice versa.

It is important to note that we cannot make direct comparisons of the variance components σ_{z00}^2, σ_{z11}^2, and σ_{z01}^2. Since the variance estimates reflect the different metrics of the variables being analyzed, simply by changing the scale of one or another variable, we could drastically alter the variance or covariance estimate. For example, by expanding the scale of a variable, say from (0,1) to (0,10), it would inflate the values of each variance estimate without changing the substantive interpretation of the results.

Note on Explained Variance

In contrast to the variance components model and the random intercept model, we are no longer able to compute estimates of explained variance in the random coefficient model. The reason for this is the distribution of the residuals is heteroskedastic (see Chap. 17) for all individual observations included in the analysis. Specifically, the total residual variance is now proportional to the value of the independent variable with the random effect. For example, if we use the equation above, the total residual will be based on the random effect for the intercept (ζ_{0j}), the individual (level 1) residual (ε_{ij}), and the product of X_{1ij} and the random effect for X_1 (ζ_{1j}):

$$\zeta_{0j} + \zeta_{1j} X_{1ij} + \epsilon_{ij}.$$

It is the product of X_{1ij} and ζ_{1j} that creates the heteroskedastic error variance, since the total residual will depend on the magnitude of X_{1ij}. In short, as X_{1ij} increases or decreases, the magnitude of the residual will change.

Bail Decision-Making Study

To gain an appreciation of the random coefficient model, we begin our analysis of the bail decision-making data by estimating regression equations for each of the 20 judges included in the sample. We continue to rely on the same simple model of logged bail as the dependent variable and number of prior drug offenses as the only independent variable in our model:

$$\log(BAIL)_{ij} = b_{0j} + b_{1j}(DRUGOFF_{1ij}) + \epsilon_{ij}.$$

Note from the subscripts to the intercept (b_0) and the slope coefficient (b_1) that there will be unique estimates for each value across the judges/clusters (*j*). Rather than present the results from 20 regression analyses in a table, the results are presented graphically in Fig. 20.3, where each line represents the regression line for a single judge. Clearly, the results show variation in the intercept across judge—note the vertical placement of each regression line that reflects larger or smaller values of the judge-specific intercept. We also note that there is variation in the regression slopes across the 20 judges—some of the slopes are steeper, some are flatter.

One way of starting to assess how similar or different the intercepts and slope coefficients are for each judge can be viewed in a scatterplot of the intercepts and the slopes. Figure 20.4 presents these results, where each point on the graph represents one of the 20 judges included in the analysis. The spread of cases

Fig. 20.3 *Judge-specific Regression Lines*

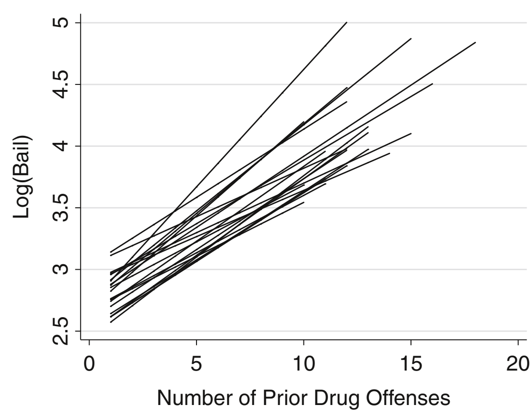
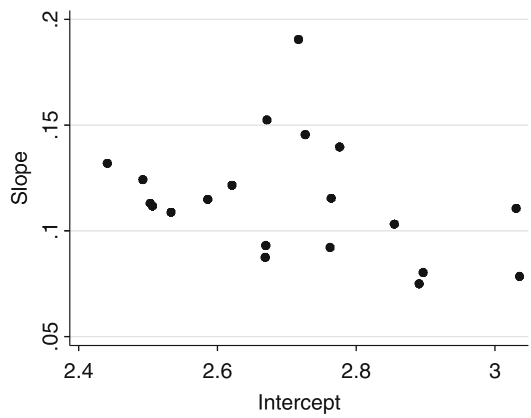


Fig. 20.4 *Scatterplot of Judge-specific Intercepts and Slope Coefficients*



across both axes confirms what we viewed in Fig. 20.3—there is variation in the intercept and the slope coefficient. Figure 20.4 shows a pattern consistent with a negative association: larger values of the intercept tend to have smaller slope coefficients, while smaller values of the intercept tend to have larger slope coefficients. The Pearson correlation for these values is -0.34. Substantively, this pattern suggests that judges with large intercepts—their cases receive relatively larger bail amounts on average—tend to place less weight on the defendant's number of prior drug offenses. In contrast, for judges with smaller intercepts—their cases tend to receive relatively lower bail amounts on average—place greater weight on the defendant's number of prior drug offenses.

The use of the random coefficient model offers a more efficient way of assessing the similarities and differences across the judges, but most importantly, will allow us to determine whether the variations in the traditional regression model are statistically meaningful or reflect random variation in the values for each judge. Table 20.7 presents the results for the random coefficient model:

$$\begin{aligned}\log(Bail)_{ij} &= (b_0 + \zeta_{0j}) + (b_1 DRUGOFF_{1ij} + \zeta_{1j} X_{1ij}) + \epsilon_{ij} \\ &= (b_0 + \zeta_{0j}) + (b_1 + \zeta_{1j}) DRUGOFF_{1ij} + \epsilon_{ij}.\end{aligned}$$

The OLS results previously reported are included in the first column of Table 20.7 to provide a ready point of comparison. Column 2 displays the results for the random coefficient model (RCM) for the 20 judges, while columns 3 and 4 present the results using grand mean centering and cluster mean centering, respectively. The estimates of the intercept and the slope coefficient in each of the columns are identical to that in the results reported in Table 20.4 for the random intercept model.

What differs in Table 20.7 is the inclusion of the random effects for the slope (σ^2_{z11}) and the covariance of the intercept and slope random effects (σ^2_{z01}). As before, we can test whether the addition of these random effects represents an improvement in the statistical model over the previous model.

Recall from above that the test of statistical significance for the addition of a random effect is a chi-square test that compares the log-likelihood values from two

Table 20.7 Regression Results for Logged Bail Amount on Number of Prior Drug Offenses

VARIABLE	OLS	RIM	RIM GRAND MEAN CENTERING	RIM CLUSTER CENTERING
Fixed Effects:				
Intercept	2.69	2.71	3.17	3.17
(se)	(0.019)	(0.037)	(0.037)	(0.041)
(*t* or *z*-score)	(141.22)	(72.95)	(85.54)	(77.72)
Number of Prior				
Drug Offenses	0.12	0.11	0.11	0.11
(se)	(0.004)	(0.006)	(0.006)	(0.006)
(*t* or *z*-score)	(29.87)	(18.37)	(18.37)	(18.27)
Random Effects:				
Intercept (σ^2_{z00})		0.0209	0.0258	0.0317
Drug Offenses (σ^2_{z11})		0.0005	0.0005	0.0005
Covariance (σ^2_{z01})		–0.0004	0.0016	0.0015

different models. In the present case, we can make two comparisons: (1) RIM v. RCM with random effect for slope coefficient and (2) RCM with random effects for the intercept and the slope v. RCM with the additional covariance of the random effects. The first comparison assesses whether the basic RCM that adds a random effect for number of drug offenses is an improvement over the RIM. The second comparison tests whether the addition of the covariance of the random effects adds to the model's improvement over and above the basic RCM.

Comparison 1:
The log-likelihood for the RIM model is -1400.57 and the log-likelihood for the RCM is -1391.74.

$$\chi^2 = 2((-1400.06) - (-1391.74)) = 16.64$$

Based on 1 degree of freedom, we find the critical χ^2, assuming a p-value < 0.05, to be 3.841. Our computed χ^2 of 16.64 has a p-value much less than 0.05, even before dividing it by 2. This result indicates that the RCM with a random effect for the slope coefficient (number of prior drug offenses) represents a significant improvement in the statistical model.

Comparison 2:
The log-likelihood for the RCM model is -1391.74 and the log-likelihood for the RCM with the covariance of the random effects is -1391.67.

$$\chi^2 = 2((-1391.74) - (-1391.67)) = 0.14$$

Based on 1 degree of freedom, we find the critical χ^2, assuming a p-value < 0.05, to be 3.841. Our computed χ^2 of 0.14 has a p-value much greater than 0.05, regardless of whether it is divided by 2. This result indicates that the RCM with the covariance of the random effects does not improve the statistical model and could be dropped from the analysis.

What do the results of these two comparisons mean? By finding the RCM makes a significant improvement in the statistical model, we know that the intercepts and the effects of number of prior drug offenses vary across the 20 judges in making bail decisions. The finding that including the covariances of the random effects does not improve the model means the judge-specific intercepts and coefficients for number of prior drug offenses are not correlated with each other. In contrast to preliminary evidence in Fig. 20.4 of a negative correlation between judge-specific intercepts and coefficients, there was no statistical evidence of such a relationship once we more formally tested the model.

Adding Cluster (Level 2) Characteristics

Thus far in our discussion of multilevel models, we have focused strictly on characteristics of the individual observations in the data—the level 1 characteristics. One of the great strengths of multilevel models is the ability to include cluster-level characteristics that will indicate how the effects of the independent variables may vary across levels of a cluster characteristic. In the example we have used thus far regarding judges and bail decision-making, we might hypothesize that characteristics of judges would affect how each would weigh information about defendants in making bail decisions. For example, gender of judge may alter the relationship that we have observed between number of prior drug offenses and bail amount. Or, years of service as a judge may affect the observed relationship between number of prior drug offenses and bail amount. These are the kinds of questions to which we now turn.

When considering adding cluster-level characteristics to a multilevel analysis, the researcher is confronted with two important questions about a cluster characteristic:

1. Is there an expectation that the cluster characteristic will **directly** affect the dependent variable?
2. Is there an expectation that the effect of an independent variable will vary by the level of the cluster characteristic?

Both of these questions force us to consider prior theory and research in thoughtfully developing our multilevel model. The first question is the more straightforward of the two questions and will often be a reflection of prior research showing that the cluster characteristic is likely important to the dependent variable being analyzed. For example, there is research indicating that gender of judge affects how criminal defendants and offenders are treated. We would have justification for hypothesizing that gender of judge would affect bail amount. Similarly, if we were studying fear of crime across a large sample of different neighborhoods, we would have justification for hypothesizing that official crime rates in the neighborhoods may have a direct affect on an individual's fear of crime.

The second question requires considerable care in developing the model, especially since there is likely to be less evidence and/or theory on which to base a hypothesis of an independent variable's (level 1) effect varying by the level of the cluster characteristic. In the literature on multilevel models, this kind of relationship is often referred to as a **cross-level interaction**, since they imply the effect of one variable (the level 1 independent variable) changes across the levels of another variable (the level 2 cluster characteristic). For example, in considering fear of crime, we may hypothesize that official crime rates may interact with the effect of age of a resident on fear of crime. Older individuals are more fearful of being crime victims in general, and we could hypothesize that as the neighborhood crime rate increased, there was a multiplier effect on the fear of crime among elderly residents. At the same time, neighborhood crime rates may not affect the level of fear of younger individuals.

Although the interpretation of cluster-level characteristics in a multilevel model may become complicated, their inclusion in the statistical model is not complicated. For the situation where we expect the cluster characteristic to have a direct effect on the dependent variable, we simply include it as an additional independent variable (denoted with a *W* in the following discussion) in our random intercept or random coefficient model. In the form of a simple random coefficient model, we would estimate the following model:

$$Y_{ij} = b_0 + b_1 X_{1ij} + b_3 W_{1j} + \zeta_{0j} + \zeta_{1j} X_{1ij} + \epsilon_{ij}. \qquad \textbf{Equation 20.14}$$

Note that our cluster characteristic W_{1j} has only a *j* subscript, indicating that the values of W_1 vary by cluster *j* but will be the same for all cases within that cluster. Like any other variable we might include in a linear regression model, the cluster characteristic can be a dummy variable or an interval level variable. The interpretation of the cluster characteristic's effect (b_3) is no different than that for other independent variables included in the model: a unit change in W_1 is expected to change the value of the dependent variable by b_3.

If we expect the effect of one of our level 1 independent variables to vary by level of a cluster characteristic, we include an interaction term between the two variables—our cross-level interaction—and add it to the model. Continuing the same random coefficient model from above, we would include an interaction between X_1 and W_1:

$$Y_{ij} = b_0 + b_1 X_{1ij} + b_3 W_{1j} + b_4 X_{1ij} W_{1j} + \zeta_{0j} + \zeta_{1j} X_{1ij} + \epsilon_{ij}. \qquad \textbf{Equation 20.15}$$

How do you interpret these results and make sense of the interaction effect? In Chap. 17, we noted that interaction effects can usually be interpreted in two different ways: we fix the value of one variable and assess the effect of the second variable. In a multilevel model, we will always fix the level (value) of the cluster characteristic first and then interpret the effect of the level 1 independent variable. In the equation above, the effect of X_1 can be written as

$$b_1 X_{1ij} + b_4 X_{1ij} W_{1j}.$$

If W_1 is a dummy variable, then for $W_1 = 0$, we have

$$b_1 X_{1ij} + b_4 X_{1ij}(0) = b_1 X_{1ij},$$

meaning the effect of X_1 at $W_1 = 0$ is simply b_1. In contrast, if $W_1 = 1$, we have

$$b_1 X_{1ij} + b_4 X_{1ij}(1) = b_1 X_{1ij} + b_4 X_{1ij} = (b_1 + b_4) X_{1ij},$$

meaning the effect of X_1 at $W_1 = 1$ is $b_1 + b_4$. For cluster-level characteristics measured on an interval scale of measurement, we would typically pick out some meaningful values and highlight the effect of X_1 at those values.

A Substantive Example: Race and Sentencing Across Pennsylvania Counties
In an analysis of sentencing decisions in Pennsylvania in the 1990s, Britt used a multilevel model to assess the effects of various social, economic, and crime measures on punishment severity decisions for offenders[8]. Of particular interest in Britt's analysis was the effect of these kinds of community characteristics on the effect of offender's race on punishment severity. For example, were black offenders punished more severely in those counties with higher rates of crime? Alternatively, were black offenders punished less severely in those counties with proportionally larger black populations? The theoretical rationale for these different hypotheses is presented in the original paper.

In what follows, we highlight a few of his key findings as examples of the power of a multilevel model. The data in the analysis represented more than 70,000 sentence length decisions spanning four years for all 67 of Pennsylvania's counties. In the context of a multilevel model, the sentence length decisions represent the individual-level data (i.e., level 1) and the counties in which the sentences were given represent the cluster-level data (level 2). We focus our discussion on the following variables:

Individual-level (Level 1)

- Sentence length: Months sentenced to jail or prison.
- Black: Coded as 1 if offender was black, 0 if offender was white (all other cases were excluded from analysis).

County-level (Level 2)

- Percentage of population classified as black.
- Percentage of population living in an urban area.
- Trend in unemployment rate (increasing, decreasing, or flat).
- Average violent crime rate.

Since our intention here is to illustrate the use and interpretation of multilevel models with cluster-level characteristics, we report abridged results in Table 20.8, showing only those elements focused on the effect of county characteristics on overall sentence length and county characteristics interacting with race of the offender. Omitted from the table are numerous case and offender characteristics relevant to predicting punishment severity, such as offense severity, criminal history, plea bargaining, and the like.

To begin the interpretation of the results, note that percentage of the population classified as black, difference in white and black per capita income, percentage living in urban areas, and trend in unemployment all have direct effects on

[8] Britt, C.L. (2000). Social context and racial disparities in punishment decisions. *Justice Quarterly*, 17, 801–826.

Table 20.8 Multilevel Regression Results for Sentence Length

VARIABLE	ESTIMATE	se	z-SCORE
Intercept	13.968	0.574	24.334
Percentage Black	–0.161	0.021	–7.750
Percentage Urban	0.026	0.007	3.910
Trend in Unemployment	0.820	0.228	3.596
Black	–2.277	0.618	–3.684
Black × Percentage Black	–0.315	0.092	–3.292
Black × Violent Crime Rate	0.009	0.003	3.000

sentence length. The other results in Table 20.8 show the direct effect for being a black offender and the interaction terms for black offender with percentage of the population classified as black and the average violent crime rate. The interpretation of the results at this point is no different than the interpretations we made in the linear regression model. For the direct effects on sentence length:

- As the percentage of blacks in a county increases, the average sentence length decreases.
- As the percentage of a county's population living in an urban area increases, the average sentence length increases.
- As the county-level unemployment rate increased over time, the average sentence length increases.

For the cross-level interaction effects of offender race with percentage of black residents in a county and average crime rate, the interpretations follow the logic to any other interaction effect. In general, what we find is that as the percentage of a county's population classified as black increases, the *effect* of being black decreases, meaning that black offenders received significantly shorter sentences than white offenders overall, but the magnitude of this difference increases as the percentage of blacks in a county increases. Conversely, in counties where the average violent crime rate was higher, the *effect* of being black *increases*, meaning that the punishments received by black offenders were more severe than those for white offenders in counties with higher violent crime rates.

The following hypotheticals will help to illustrate these patterns. Suppose that we have four different counties with the following characteristics:

- County A: Percentage Black = 10, Violent Crime Rate = 100
- County B: Percentage Black = 20, Violent Crime Rate = 100
- County C: Percentage Black = 10, Violent Crime Rate = 200
- County D: Percentage Black = 20, Violent Crime Rate = 200

The equation for the effect of being a black offender on sentence length is:

$$-2.277\,Black - 0.315\,Black \times PercentBlack + 0.009\,Black \times ViolentCrimeRate.$$

The effect of being a black offender in County A:

$$-2.277\,Black - 0.315 \times 10 \times Black + 0.009 \times 100 \times Black = -4.527\,Black.$$

The effect of being a black offender in County B:

$$-2.277\,Black - 0.315 \times 20 \times Black + 0.009 \times 100 \times Black = -7.677\,Black.$$

The effect of being a black offender in County C:

$$-2.277\,Black - 0.315 \times 10 \times Black + 0.009 \times 200 \times Black = -3.627\,Black.$$

The effect of being a black offender in County D:

$$-2.277\,Black - 0.315 \times 20 \times Black + 0.009 \times 200 \times Black = -6.777\,Black.$$

In all cases, the effect of being a black offender resulted in a shorter sentence, ranging from about 3.6 months to just under 7.7 months.[9] For the two pairs counties with matching percentage black populations, the increase in the violent crime rate resulted in a shrinking of the effect of being black and moving the coefficient closer to 0, where there is no difference between black and white offenders. For the two pairs of counties with matching violent crime rates, as the percentage of the black population increased, the sentence disparity increased further, with black offenders receiving even more lenient sentences.

Chapter Summary

Multilevel models offer an important extension to traditional linear regression models by statistically accounting for possible clustering in a sample of data. Observations that come from the same cluster (e.g., multiple survey respondents within the same neighborhood, multiple cases processed by a judge or prosector, and so on) will tend to be more similar to each other than to observations from different clusters. This results in an increased likelihood of finding statistically significant effects, since many of the cases within a cluster will exhibit a similar pattern of association. Multilevel models account for clustering by allowing for random variation in the intercepts and possibly the coefficients of the independent variables. Models that allow for variation in the model intercept are referred to as **random intercept models (RIMs)**, while models that contain a random intercept and at least one random slope coefficient are referred to as **random coefficient models (RCMs)**.

Variation in both the intercept and the coefficient for an independent variable are measured with what are called **variance components**—measures of how much the intercept and slope may vary across cluster. These are also called

[9] While this finding often strikes many criminal justice students as counterintuitive, it is consistent with much of the research done on the race effects on sentencing outcomes. What is not shown here, but is included in Britt's article, is the effect of race on the likelihood of being incarcerated, where black offenders were much more likely than white offenders to be sentenced to prison. The findings here just highlight that once sentenced to incarceration, the length of time is shorter for black offenders compared to white offenders. In Pennsylvania, this has typically taken the form of more black offenders being sentenced to local jails for relatively short periods of time, while white offenders who have received incarceration sentences will be sent to state prisons for relatively longer stays.

the **random effects**. We test the significance of the variance estimates with a chi-square test that compares the model with the random effects against a linear regression model without any random effects.

In estimating multilevel models, we may also center the values of the independent variables. Centering can take on two forms: grand-mean or cluster-mean centering. Centering has no effect on the interpretation of the slope coefficients, but will alter the substantive meaning of the model intercept. In **grand-mean centering**, the model intercept represents the overall sample mean for the dependent variable. In **cluster-mean centering**, the model intercept represents a weighted sample mean for the dependent variable that is conditioned on the number of cases per cluster. The more balanced the size of the clusters, the more similar the two estimates of the model intercept will be. In general, centering the values of the independent variables will tend to simplify the estimation of the overall multilevel model.

Centering also allows for the testing of different **between** and **within group effects** of the independent variables. Much of the research in criminology and criminal justice assumes the between and within effects are the same without ever testing for similarity. By estimating models that include the cluster means as independent variables, it is possible to assess directly how or whether the between cluster effects are the same as the within cluster effects. If the results of these tests indicate the effects are the same, then grand-mean centering is appropriate. Alternatively, if the between and within effects are different, then cluster-mean centering is a more appropriate technique.

Key Terms

between effect Effect of an independent variable on the dependent variable using the cluster as the unit of analysis—a regression of cluster-level averages across all the clusters included in the analysis.

cluster-mean centering Computed difference between the observed raw score on some variable for each observation in the sample and the cluster mean for that variable.

contrast coding A method for recoding a multi-category nominal variable into multiple indicator variables (one less than the total number of categories), where the indicator category is coded as 1, the reference category is coded as − 1, and all other categories are coded as 0. Contrast coding ensures that the sum of all the estimated effects for the indicator variable is equal to 0.

cross-level interaction An interaction effect included in a multilevel model between a level 1 independent variable and a level 2 cluster characteristic.

fixed effects A descriptive label for the regression coefficients (b_k) estimated in a model with random effects. Fixed effects represent the average effects of the independent variables on the dependent variable across all individuals and clusters in a multilevel model.

grand-mean centering Computed difference between the observed raw score on some variable for each observation in the sample and the overall sample mean for that variable.

intraclass correlation A measure of association that measures the level of absolute agreement of values within each cluster.

multilevel data Sample data where individual observations (level 1 data) are clustered within a higher-level sampling unit (level 2 data).

random coefficient model A linear regression model that allows the intercept and the effect of at least one independent variable to vary randomly across cluster—random effects are included for the model intercept and at least one independent variable.

random effects A descriptive label for the random error terms included in a multilevel model that allow for variation across cluster from the sample average estimated in the fixed effects. Random effects are assumed to be normally distributed in most multilevel models.

random intercept model A linear regression model that allows the intercept to vary randomly across cluster—random effects are included for the model intercept.

regression coding A method for recoding a multi-category nominal variable into multiple indicator dummy variables (one less than the total number of categories), where the indicator category is coded as 1 and all other categories are coded as 0. The reference category does not have an indicator variable and is coded as a 0 on all the indicator dummy variables.

variance components model A one-way analysis of variance model that includes random effects for each cluster that assesses whether there is random variation in the mean of the dependent variable across the clusters included in the analysis.

within effect Effect of an independent variable on the dependent variable within each cluster and then averaged across all clusters or groups included in the analysis.

Symbols and Formulas

ζ_{0j} Random effect for the model intercept b_0

ζ_{kj} Random effect for the regression coefficient for the independent variable k (bk)

σ^2_{z0} variance of the random effect for the model intercept in a variance components model

σ^2_{z1} variance of the random effect for the model intercept in a random intercept model

σ^2_{e0} error variance (unexplained variance) in a variance components model

σ^2_{z1} error variance (unexplained variance) in a random intercept model

σ^2_{z00} Variance of the random effect for the model intercept in a random coefficient model

σ^2_{zkk} Variance of the random effect for the independent variable k in a random coefficient model

σ^2_{z0k} Covariance of the random effects for the model intercept and for the independent variable k in a random coefficient model

General equation for the variance components model:

$$Y_{ij} = b_0 + \zeta_j + \epsilon_{ij}$$

General equation for the random intercept model with one independent variable:

$$Y_{ij} = b_0 + b_1 X_{1ij} + \zeta_{0j} + \epsilon_{ij}$$

General equation for the random coefficient model with one independent variable:

$$Y_{ij} = (b_0 + \zeta_{0j}) + (b_1 + \zeta_{1j})X_{1ij} + \epsilon_{ij}$$

Likelihood-ratio test for variance components:

$$\chi^2 = -2(LL(\mathbf{Model}1) - LL(\mathbf{Model}2))$$

Equation for computing the intraclass correlation (ρ):

$$\rho = \frac{\sigma_z^2}{(\sigma_z^2 + \sigma_e^2)},$$

Equation for grand-mean centering:

$$X_{ij} - \bar{X}_{..}$$

Equation for cluster-mean centering:

$$X_{ij} - \bar{X}_{.j}$$

Equation for explained variance in a random intercept model:

$$R^2 = \frac{(\sigma_{z0}^2 + \sigma_{e0}^2) - (\sigma_{z1}^2 + \sigma_{e1}^2)}{\sigma_{z0}^2 + \sigma_{e0}^2}$$

Equation for the explained variance at level 2 of a random intercept model:

$$R_z^2 = \frac{\sigma_{z0}^2 - \sigma_{z1}^2}{\sigma_{z0}^2}$$

Equation for the explained variance at level 1 of a random intercept model:

$$R_e^2 = \frac{\sigma_{e0}^2 - \sigma_{e1}^2}{\sigma_{e0}^2}$$

Exercises

1. Researchers interested in the possible effects of neighborhood poverty on patterns of intimate partner violence (IPV) gathered interview data from 100 residents in each of a city's 53 neighborhoods. As a first step in establishing neighborhood variability in IPV, the researchers estimated a variance components model and obtained the following results:

   $$-2LL(\mathbf{ANOVA}) = -2000$$

   $$-2LL(\mathbf{REM}) = -1900$$

$\sigma_z^2 = 0.10$

$\sigma_e^2 = 0.40$

a. Test whether there is significant variation in the prevalence of IPV across these 53 neighborhoods. Explain the substantive meaning of your findings.

b. To what extent does neighborhood affect the prevalence of IPV? Calculate the intraclass correlation coefficient and explain its substantive meaning.

2. In a study of 13,726 students across 491 schools, a study asked about self-reported delinquent behavior, which was measured with a scale that ranged from 0 (no delinquency) to 7 (high rate of delinquency). The researchers were particularly interested in the effects of academic performance on delinquent behavior and estimated a random intercept model, obtaining the following results:

VARIABLE	COEFFICIENT
Intercept	0.50
GPA	− 0.30
Educational Aspirations	− 0.05
Father's Education	− 0.20
Mother's Education	− 0.35
σ_z^2	0.15
σ_e^2	0.95

a. Calculate the intraclass correlation coefficient for the variance components model and explain its substantive meaning.

b. Calculate the intraclass correlation coefficient for the random intercept model and explain its substantive meaning.

c. Interpret the change in the value of the intraclass correlation coefficient between the variance components and random intercept models.

d. How would the intraclass correlation coefficient change if additional covariates were added to the model? Describe how the statistical significance of the added covariate affects the change in the value of the intraclass correlation coefficient in this case.

3. A study of anti-social behavior among children collected information on 674 families, each with at least two children who could participate in the study. Based on prior research, the investigators expected the within-family and between-family effects of parental attachment to be different. The investigators found the following effects:

$b_{Attachment} = -0.23, se = 0.07$

$b_{ClusterMeanofAttachment} = -0.16, se = 0.03$

a. Explain whether the investigators found evidence of different within-family and between-family effects of parental attachment.

b. Which type of centering would be most appropriate for these data if the investigators simply want to estimate a single effect for parental attachment that ignores the between and the within effects?

4. Researchers interested in studying the effects of neighborhood characteristics on individuals' perceptions of fear of crime victimization selected a random sample of 100 neighborhoods and then interviewed 50 residents within each neighborhood. Using a fear of crime scale as the dependent variable and demographic characteristics as covariates, the researchers estimated a series of regression models to test for random effects across neighborhood. The results appear in the following table (assume all fixed effects are statistically significant with $p < 0.05$):

VARIABLE	OLS	RANDOM INTERCEPT MODEL	RANDOM COEFFICIENT MODEL
Fixed Effects:			
Intercept	4.50	4.40	4.50
Age	0.07	0.06	0.05
Female	0.78	0.80	0.90
Black	0.64	0.52	0.48
Random Effects:			
Intercept (σ^2_{00})		0.03	0.03
Age (σ^2_{11})			0.01
Female (σ^2_{22})			0.12
Black (σ^2_{33})			0.05
Model Information:			
– 2L L	–3176	–2273	–2095

a. Test whether there is significant variation in the model intercept across the 100 neighborhoods using a significance level of 5%. Interpret your result.

b. Test whether there is significant variation in the coefficients for age, female, and black using a significance level of 5%. Interpret your result.

5. A study of sentence length decisions began by selecting a random sample of 35 judges within a large state. For each judge, the researchers selected a random sample of 200 cases among those involving sentences to jail or prison. Sentence length was measured as the number of months sentenced to incarceration. To assess the impact of legal characteristics on sentence length decisions, the researchers developed measures of severity of the conviction crime and of criminal history. After establishing that the intercept varied across judge, the researchers investigated a series of random coefficient models that examined whether there were correlations of the random effects for the intercept and the two covariates. The following table presents their results:

VARIABLE	RIM	RCM 1	RCM 2
Fixed Effects:			
Intercept	5.19	5.23	7.98
Severity of Offense	2.72	2.65	1.95
Criminal History	5.31	5.62	4.97
Random Effects:			
Intercept (σ^2_{00})	0.29	0.26	0.21
Severity of Offense (σ^2_{11})		0.16	0.12
Criminal History (σ^2_{22})		0.21	0.18
Intercept-Severity (σ^2_{01})			0.06
Intercept-History (σ^2_{02})			0.04
Severity-History (σ^2_{12})			0.11
Model Information:			
$-2LL$	–2984	–2781	–2779

a. Test whether there is significant variation in the coefficients for offense severity and criminal history using a significance level of 5%. Interpret your result.

b. Test whether the addition of the random effect covariances is statistically significant using a significance level of 5%. Interpret your result.

Computer Exercises

We have noted throughout this chapter that multilevel models can become complex very quickly and so we have tried to keep our focus on the basic elements of multilevel models. The essential syntax required to estimate these models is generally straightforward and does not become complicated until we start customizing the statistical model. The examples below and in the accompanying syntax files for SPSS (Chapter_20.sps) and Stata (Chapter_20.do) illustrate key components to the multilevel commands without getting bogged down in too many of the options and details.

SPSS

In SPSS, random intercept models and random coefficient models are both estimated with the MIXED command:

```
MIXED dep_var WITH list_of_indep_vars
/PRINT = SOLUTION TESTCOV
/FIXED = INTERCEPT list_of_indep_vars
/RANDOM = INTERCEPT random_indep_var(s) |
SUBJECT(cluster_variable) COVTYPE(UN) .
```

where the first line has the same structure as many of the other commands in SPSS. The /PRINT=SOLUTION TESTCOV option requests SPSS to print out the coefficient table (SOLUTION) and to test the random effect variances and covariances for statistical significance (TESTCOV). The /METHOD=ML option forces SPSS to estimate the models with maximum likelihood, while the FIXED = option line should include INTERCEPT and all of the independent variables included in the model. The /RANDOM = option will then determine whether a RIM or RCM will be estimated. If a RIM is to be estimated, the /RANDOM line simplifies to:

/RANDOM = INTERCEPT | SUBJECT(cluster_variable) .

For an RCM that only estimates random effects variances:

/RANDOM = INTERCEPT random_indep_var(s) |

SUBJECT(cluster_variable) .

For an RCM that estimates random effects covariances and variances you will need to add the COVTYPE(UN) option to the /RANDOM line, since the default in SPSS is to estimate an RCM without random effect covariances:

/RANDOM = INTERCEPT random_indep_var(s) |

SUBJECT(cluster_variable) COVTYPE(UN) .

The output from running any of these commands is the same and includes tables of coefficients (the fixed effects), of covariances and variances for the random effects, and of model summary statistics.

Stata

Random Intercept Models

To estimate random coefficient models in Stata, you will have access to two different commands: **xtreg** and **xtmixed**. The structure to the xtreg command is:

xtreg depvar list_of_indep_vars, **i**(cluster_variable) **mle var**

The basic format is similar to the regress command—the difference follows the comma, where the **i()** option indicates which variable provides information on the cluster. In this chapter, the cluster was the judge identifier. The **var** option is included to force Stata to estimate the variance of the random intercept—the default output in Stata is to report the square root of the variance (i.e., the standard deviation of the random effect). Finally, the **mle** option forces Stata to compute the maximum likelihood estimates for the RIM.

The use of the **xtmixed** command is similar:

xtmixed depvar list_of_indep_vars **||** cluster_variable:,
mle var

where instead of a comma immediately following the list of independent variables, we have two vertical bars (||) that are followed by the cluster variable with a colon (**:**) appended, and then the comma and request for maximum likelihood estimates. As we explain in the next section regarding the estimation of random coefficient models, the vertical bars will be useful for designating random coefficients.

Random Coefficient Models

The **xtreg** command cannot be used for random coefficient models, while **xtmixed** can be used. To use **xtmixed** for an RCM, we simply add the name of the independent variable(s) after the **:** that we want to estimate random effects for:

xtmixed depvar list_of_indep_vars || cluster_variable: random_indep_var(s), **mle var**

The structure to the rest of the command is the same—all we do is include one or more independent variable names after the **:** and before the comma. Note that the default RCM in Stata is to estimate a model with no covariances of the random effects. If we want to estimate the covariances of the random effects, we add the option **cov(unstructured)** to the command line:

xtmixed depvar list_of_indep_vars || cluster_variable: random_indep_var(s), **mle var cov(unstructured)**

The output will then contain the variances for the intercept, any independent variables with effects allowed to vary across cluster, and all possible covariances of the random effects.

Problems

1. Open the Bail Decision-Making data file (bail-data-example.sav or bail-data-example.dta). The sample syntax files for this chapter include the syntax required to reproduce most of the tables in this chapter. Work your way through one of the syntax files and make sure you understand how it works.
2. Using the Bail Decision-Making data file, use logbail as the dependent variable. Select a set of 4–6 independent variables and estimate the following models:
   a. Random intercept model:
      i. Which of the fixed effects are statistically significant at a 5% level of significance?
      ii. Interpret each of the statistically significant fixed effects.
      iii. Test whether the RIM offers a significant improvement over a linear regression model. Use a significance level of 5%.
   b. Random coefficient model: Select at least one, but no more than three, of the independent variables to have a random coefficient. Do not estimate the covariances of the random effects.
      i. Did anything change in regard to the effects of these independent variables? Value of the coefficient? Level of statistical significance?
      ii. Test whether this RCM offers a significant improvement over the RIM estimated in part (a). Use a significance level of 5%.
   c. Random coefficient model: Use the same RCM as in part (b) but estimate the random effect covariances.
      i. Did anything change in regard to the effects of these independent variables? Value of the coefficient? Level of statistical significance?
      ii. Test whether this RCM offers a significant improvement over the RCM estimated in part (b). Use a significance level of 5%.

3. Continue to use the Bail Decision-Making data file, change the dependent variable to the bail amount requested and estimate the same set of models with the same independent variables as in Question 2.

   a. Explain how the results are similar. Different. Focus on the values of the coefficients and the covariances and variances of the random effects.

   b. What might account for these differences? (Hint: You may want to generate histograms for both dependent variables as a starting point.)