14

Random Coefficient Regression and Multilevel Models

OLS regression and the regression approaches subsumed under the generalized linear model, including logistic and Poisson regression, all assume that observations are independent of one another. If observations on two individuals are independent of one another, then knowledge of scores on one individual provides no information whatever about scores on the other individual. Put another way, there is no relationship between the measures on one individual and measures on any other individual. Should there be repeated observations on the same individuals, standard OLS regression analyses assume that these observations are independent across time.

14.1 CLUSTERING WITHIN DATA SETS

Our observations may well not be independent. Sections 4.3.1 and 4.4.5 introduced the issue of dependency among residuals, which may occur when the cases are members of an intact group, (e.g., a family, a community organization). Consider, for example, the IQ scores of children within families. We expect correlation among these IQ scores; the IQ scores of children within a single family may be more similar to one another in value than would be expected in a random sample of children. Measures taken on members of dyads—for example, spouses or twins—are also highly interrelated. Dependency can arise in measures of demographic characteristics of individuals who live in close proximity (e.g., among the incomes or ethnicities of individuals who live in particular neighborhoods of a large metropolitan area). Dependency can also occur in experimental research when we run participants in groups, and the behavior of group members can influence the responses of individuals, for example, in experiments in group processes in social psychology. Under some circumstances dependency can arise in experiments when the presentation of the treatment condition inadvertently varies slightly from session to session in which groups of subjects participate. Correlation or dependency among subsets of cases within a data set, as reflected in all these examples, is referred to as *clustering*.

Dependency in data can also arise when we take repeated measures on single individuals over time. For example, we might measure the anxiety level of each of a set of individuals once a month for a period of months. We would expect that the anxiety measures from any one individual would be more correlated with one another than the anxiety measures across individuals. This is another form of dependency in data, referred to as *serial dependency* (see Section 4.4.5). In this chapter we focus on clustering among individuals within groups, and approaches to handling data that contain such clustering of individuals. Chapter 15 is devoted to the treatment of repeated measures data that may exhibit serial dependency. Much of what is developed here for the treatment of clustering among individuals generalizes directly to serial dependency; Section 15.4 develops this generalization.

When data are clustered, OLS regression may lead to inaccuracies in inference. The *random coefficient regression model*, an alternative to OLS regression, is structured to handle clustered data. The random coefficient (RC) regression model differs from OLS regression in the assumptions made about the nature of the regression coefficients and correlational structure of the individual observations. Further, when individuals are clustered into groups, we may have multiple levels of measurement, at both the individual and the group level. For clients in therapy groups, for example, we may measure characteristics of the clients (individual level) and characteristics of the therapist that impact all members of the group (group level). Measures taken on multiple levels may be treated in *multilevel models* (equivalently termed *hierarchical linear models*), which employ random coefficient regression (Goldstein, 1995; Kreft & de Leeuw, 1998; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999).

14.1.1 Clustering, Alpha Inflation, and the Intraclass Correlation

As indicated, clustering poses difficulties for statistical inference in the general linear model and generalized linear model frameworks. If data are clustered, the standard errors of OLS regression coefficients are typically negatively biased (i.e., too small). Thus the confidence intervals around individual regression coefficients are typically too small. Statistical tests for the significance of individual regression coefficients, which involve division of the coefficients by their standard errors, will in general be too large, leading to overestimation of significance, or *alpha inflation*. As clustering increases (i.e., as scores within clusters become increasingly similar to one another), alpha inflation increases as well; that is, the actual level of Type I error increasingly exceeds the nominal level. This same bias occurs if logistic regression is applied to clustered data.

The degree of clustering, (i.e., the degree of correlation or nonindependence among a set of observations), is measured by the *intraclass correlation* (ICC; Shrout & Fleiss, 1979). The ICC measures the proportion of the total variance of a variable that is accounted for by the clustering (group membership) of the cases. The ICC also can be conceptualized as a measure of the extent to which members of the same category (for example, children within families) are more similar to one another than to members of other categories.[1] Put another way, the ICC measures whether scores from different groups are more discrepant from one another than scores within the same group. The ICC ranges from 0 for complete independence of observations to 1 for complete dependence.[2] An assumption underlying the general linear model and generalized linear model is that ICC = 0.

---

[1]There are multiple definitions of the intraclass correlation (ICC) that arise in the context of varying experimental designs. The ICC is often used as a measure of interjudge reliability in designs in which multiple judges rate multiple targets. Shrout and Fleiss (1979) provide an explication of the multiple definitions of the ICC in the estimation of interjudge reliability.

[2]This presentation of the ICC assumes that clustering produces similarity or positive correlation among cases within a cluster; this is by far the usual situation in clustered data. In rare instances, for a particular research purpose, an experimenter may create clusters in which the individuals are sampled to be highly discrepant from one another, (i.e., more dissimilar from one another than might be expected by chance alone). In this exceptional instance, the ICC can be negative and lead to an actual Type I error rate lower than the nominal Type I error rate (Kenny & Judd, 1986).

14.1.2 Estimating the Intraclass Correlation

We can calculate the ICC for any variable in a data set with a clustered structure, that is, in which individuals are members of groups or clusters. Using common notation, if we let τ represent the amount of variance in a variable that is due to differences among groups in the population, and σ^2 represent the variance among scores within groups, pooled across all groups, then the total variance is given as $\tau + \sigma^2$. The population expression for the ICC is given as follows:[3]

$$\text{ICC} = \frac{\tau}{\tau + \sigma^2}. \tag{14.1.1}$$

If groups do not differ from one another, then $\tau = 0$, and the ICC $= 0$.

Data on a common dependent variable taken across a set of groups are required to estimate the ICC (e.g., the educational attainment of children within families, each of which has at least two children). The ICC can be estimated from a fixed effects one-factor nonrepeated measures analysis of variance (ANOVA) in which the factor is the grouping variable (for example, family), and the levels are the particular groups (the particular families). Members of the groups (children within families) serve as the observations. For those familiar with ANOVA, the estimate $\hat{\tau}$ of τ is taken from $\text{MS}_{\text{treatment}}$ and MS_{error}, where $\hat{\tau} = \text{MS}_{\text{treatment}} - \text{MS}_{\text{error}}/n$, where n is the number of cases per cell for equal group sizes. The estimate $\hat{\sigma}^2$ of σ^2 is taken directly from MS_{error} as $\hat{\sigma}^2 = \text{MS}_{\text{error}}$. Then the estimate of the ICC based on the fixed effects one-factor ANOVA is given as follows:

$$\text{ICC} = \frac{\text{MS}_{\text{treatment}} - \text{MS}_{\text{error}}}{\text{MS}_{\text{treatment}} + (n-1)\text{MS}_{\text{error}}}. \tag{14.1.2}$$

For unequal group sizes, n is replaced by $\tilde{n} = M_n - [sd^2(n_j)/(gM_n)]$, where M_n is the mean number of cases per group, $sd^2(n_j)$ is the variance of the number of cases per group, and g is the number of groups (Snijders & Bosker, 1999).

For those familiar with fixed effects one-factor ANOVA, examination of the ICC in relation to the omnibus F test for the one-factor ANOVA may provide more insight into the ICC. The omnibus F test in one-factor fixed effects ANOVA is given as $F = \text{MS}_{\text{treatment}}/\text{MS}_{\text{error}}$. We see that if $\text{MS}_{\text{treatment}} = \text{MS}_{\text{error}}$, indicating that there is no effect of the grouping factor, then the ICC $= 0$.

An ICC of .01 or .05 may seem very small; however, the actual alpha level of statistical tests increases dramatically even with such apparently small ICCs. For example, in a one-factor fixed effects ANOVA with $n = 25$ cases per cell and an ICC of only .01, the actual alpha level for the test of the treatment factor is .11 when the nominal alpha level (the alpha level for the tabled critical value) is .05. With $n = 25$ cases per cell and an ICC of .05, the actual alpha level is .19 for nominal alpha of .05. The alpha inflation increases as the ICC and sample size increase (see Barcikowski, 1981, p. 270; Kreft & de Leeuw, 1998, p. 10 for further numerical examples). This same sort of alpha inflation occurs in regression analysis with clustered data.

---

[3]This expression corresponds to ICC(1, 1) of Shrout and Fleiss (1979), one of a number of expressions for the ICC, which vary across designs and data structures.

14.2 ANALYSIS OF CLUSTERED DATA WITH ORDINARY LEAST SQUARES APPROACHES

Clustered data historically have been the cause of great statistical hand wringing and analytic acrobatics. Clustering has been viewed as a "problem" with data, something that requires handling in order to get on with the study of the relationships of predictor measures on individuals (e.g., the IQ scores of individual children within families) to some DV of interest (e.g., educational attainment, or the level of education achieved by each child). In the OLS regression context, three approaches have been taken to examining the relationship of predictors to a dependent variable when data are clustered (have group structure). The first is to ignore clustering and analyze the individual cases as if there were no group structure in the data, referred to as *disaggregated analysis*. In the IQ example, we might predict educational attainment from the child's IQ, ignoring the fact that there may well be clustering among siblings, who appear in the same data set. Here we expect alpha inflation. The second is to aggregate data at the group level, obtaining a mean on each predictor variable and on the DV for each group; the groups are then treated as the unit of analysis, referred to as *aggregated analysis*. In the IQ example, we might relate the mean IQ of the children in each family to the mean level of schooling achieved by the children of that family. There are conceptual difficulties with this approach. The resulting regression equation describes the relationship of the means of predictors in individual clusters to the mean of the dependent variable in those clusters. We set out to study how the IQ scores of individual children relate to their level of educational attainment. However, the aggregated analysis tells us how mean child IQ in a family relates to mean level of schooling in that family. Generalizing or more correctly particularizing from the group equation to the individual might lead to very inaccurate conclusions; generalization from results at one level of aggregation to another (or unit of analysis to another) is referred to as the *ecological fallacy* (Robinson, 1950).

The disaggregated analysis and the aggregated analysis estimate very different regression coefficients. Using the terminology of analysis of covariance (ANCOVA), the disaggregated analysis estimates the *total regression coefficient* of the criterion on the predictor, B_T, illustrated in Fig. 14.2.1(A). The aggregated analysis estimates the *between-class regression coefficient*, B_B. Again there is a single regression equation; here each group is treated as a single case. The disaggregated versus the aggregated analyses may yield very different results. Kreft and de Leeuw (1998) provided an example in which this is so; they examined the relationship between education (the predictor) and income (the dependent variable) within 12 industries: The total regression coefficient in the disaggregated analysis was positive, with higher education associated with higher income overall. However, the between-class coefficient was negative. Overall, the highest paid industries had lower average education levels than did lower paid industries (e.g., there was lower average education and higher average salary in the transportation industry; in the teaching industry, there was higher average education and lower average salary).

The third OLS approach is to analyze the regression of a dependent variable on predictors of interest at the individual case level, but to include as predictors a set of $(g - 1)$ dummy codes for g groups or clusters to identify the group membership of each individual in the data set (see Section 8.2). Consider Fig. 14.2.1(B), which shows a series of regression lines, one for each group. The regression lines all have the same slope; however, all the intercepts differ, indicating that the different groups have different arithmetic mean levels on the dependent variable. Having the $(g - 1)$ dummy codes as predictors takes into account these differences in intercepts of the individual groups. There is only one slope, a constant across all the groups; that is, it is assumed that the slope of the regression of Y on X is constant across all groups.

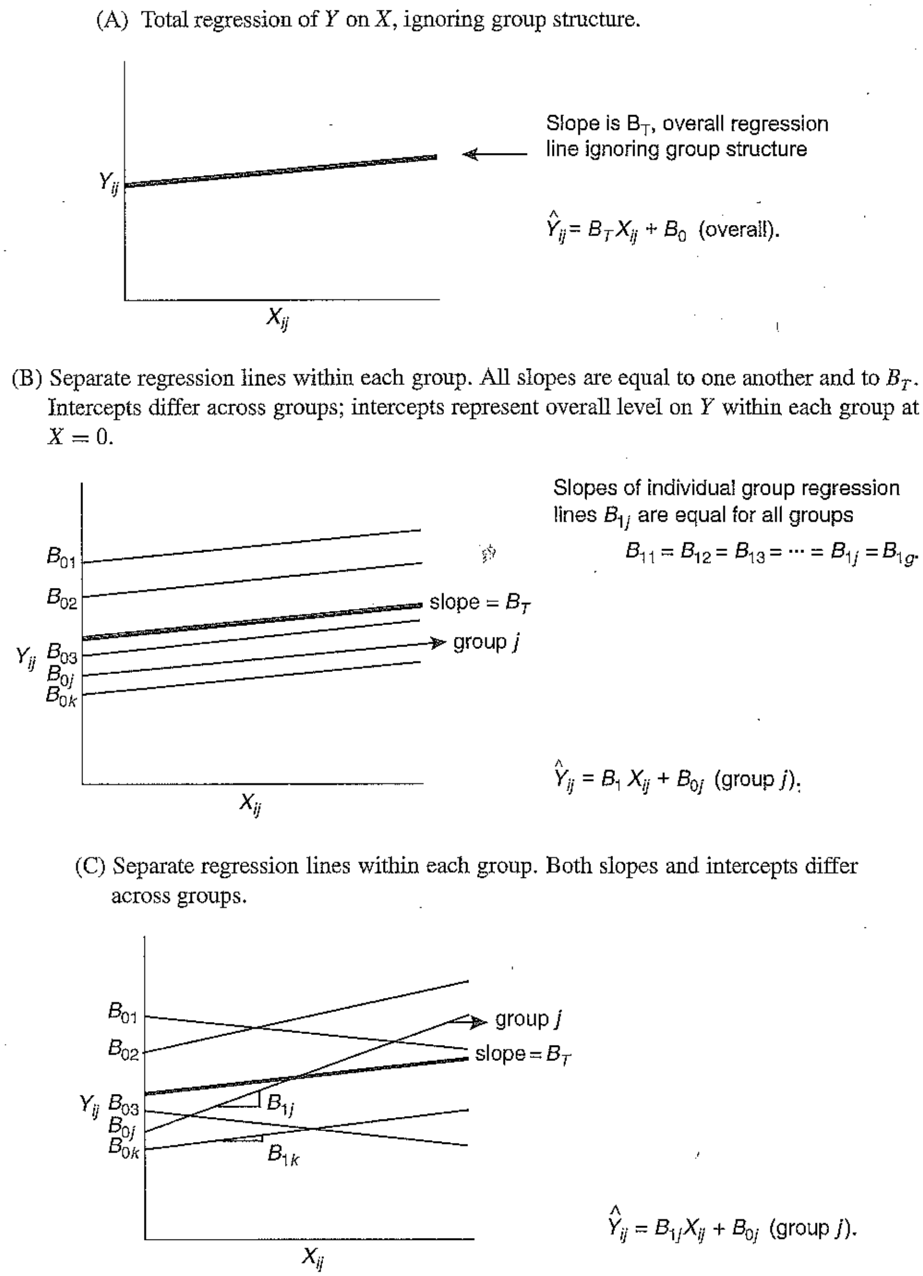


FIGURE 14.2.1 Slopes and intercepts in OLS and random coefficient models.

Once the dummy codes for the groups are included in the analysis, the regression coefficient for the predictor in question (here, child's IQ) is the *pooled within-class regression coefficient*, B_W. The pooled within-class regression coefficient is the weighted average of the regression coefficients in each of the individual groups. This approach is, in fact, the analysis of covariance (ANCOVA) model (see Section 8.7.5). Typically, when we use ANCOVA, the focus is on the effect of groups on the outcome when the individual level predictor (the covariate) is partialed out. In contrast, here the focus is on the relationship of the individual level predictor to the dependent variable when differences among group means are partialed out. This third analysis corrects for any differences in the means of the groups when the predictive utility of particular individual level predictors is assessed (e.g., the regression of educational attainment on IQ with family membership controlled).

It is also possible to model interactions between group membership and predictors; for example, one could include in the regression analysis interaction terms between each dummy code and IQ; the set of interaction terms measures whether the relationship of IQ to educational attainment varies across families. In Fig. 14.2.1(C), each group has a different intercept and a different slope. The differences in intercepts would be captured by the $(g - 1)$ dummy codes discussed earlier. Another $(g - 1)$ interaction codes, the cross product of each dummy code with the predictor in question (here, IQ), would be required to capture the differences in slopes.

The approach to handling clustering with dummy codes to account for group mean differences on the dependent variable is often referred to as the *fixed effects approach to clustering* (see, for example, Snijders & Boskers, 1999). This name arises because OLS regression analysis, on which the approach is based, is also referred to as a *fixed effects* regression analysis, for reasons explained in Section 14.4. As is explained in Section 14.14, under some conditions, (e.g., small numbers of clusters in the data set) this approach is recommended for the analysis of clustered data.

14.2.1 Numerical Example, Analysis of Clustered Data With Ordinary Least Squares Regression

In this section we present a numerical example of the use of OLS approaches to handling clustered data. Here we introduce a simulated numerical example of the prediction of weight loss as a function of motivation to lose weight; we will use this example throughout the chapter. We assume that the data have been collected from intact women's groups that have a focus on diet and weight control; the groups meet regularly to discuss diet and weight control, and have some level of cohesion. We may thus expect some correlation among the women within a group j in both their motivation to lose weight and weight loss success. There are a total of 386 women in all distributed across the 40 groups. Group size ranges from 5 to 15 women. There is substantial clustering in the data, reflected in the fact that the groups differ substantially in mean pounds of weight lost, from a low mean of 9.75 pounds lost to a high mean of 24.43 pounds lost. Using Eq. (14.1.2) we estimate the ICC. Based on a fixed effects one-factor analysis of variance with the 40 groups as levels of the group factor, the ICC is estimated at .22. Specifically, $\text{MS}_{\text{treatment}} = 60.0225, \text{MS}_{\text{error}} = 16.0231$, and $\tilde{n} = 9.63$, and the ICC $= (60.0225 - 16.0231)/[60.0225 + (9.63 - 1)(16.0231)] = .22$.

Disaggregated Analysis

Table 14.2.1 provides analyses of these data by three OLS approaches. The disaggregated analysis is given in Table 14.2.1A. Here weight loss in pounds (POUNDS) for each individual woman is predicted from motivation to lose weight (MOTIVATC); MOTIVATC is measured on a six-point scale and is *centered* around the grand mean of the 386 cases. Group membership is completely ignored. This analysis yields the total regression coefficient $B_T = 3.27$ with a standard error of .15. With $n = 386$ cases and a single predictor $(k = 1)$, there are $(n - k - 1)$ degrees of freedom for $\text{MS}_{\text{residual}}$, and (1, 384) df for the F test for overall regression. The total regression coefficient $B_T = 3.27$ indicates that on average, across all cases, with group structure ignored, there is a 3.27-pound predicted weight loss for each one unit increase in motivation.

CH14EX01

TABLE 14.2.1
OLS Regression of Pounds Lost on Motivation With Three Approaches to Clustering Within the Data

A. Disaggregated analysis of individual cases with clustering ignored
$R^2 = .545$, $F(1, 384) = 459.57$, $p < .001$

Variable	B	$SE\ B$	Beta	t	Significance of t
MOTIVATC	3.270	.153	.738	21.438	.001
(Constant)	15.003	.156		96.409	.001

B. Aggregated analysis using group means of predictor and criterion
$R^2 = .492$, $F(1, 38) = 36.80$, $p < .001$

Variable	B	$SE\ B$	Beta	t	Significance of t
MOTMEANC	4.162	.686	.701	6.067	.001
(Constant)	15.159	.304		49.910	.001

C. Disaggregated analysis with dummy coded groups ($k - 1 = 39$ dummy codes in all)
39 Group Codes: $R^2 = .297$, $F(39, 346) = 3.75$, $p < .001$
39 Group Codes plus MOTIVATC: $R^2 = .704$, $F(40, 345) = 20.48$, $p < .001$

Variable	B	$SE\ B$	Beta	t	Significance of t
GR1	−1.192	1.098	−.0419	−1.086	.278
GR2	1.254	1.130	.0419	1.110	.268
.	.	.	.	.	.
GR38	2.444	1.285	.067	1.902	.058
GR39	−2.931	1.178	−.092	−2.488	.013
MOTIVATC	3.119	.143	.704	21.765	.001
(Constant)	15.264	.722		21.138	.001

Note: The notation GRXX refers to the dummy code for group XX. For example, GR38 refers to the dummy code in which group 38 is coded "1"; all other groups, "0".

Aggregated Analysis

CH14EX02

The aggregated analysis is given in Table 14.2.1B. The dependent variable is the mean weight loss in the group; the predictor, the mean motivation level in the group. There are 40 groups and hence 40 cases in all. Mean weight loss is predicted from mean motivation level, yielding the between class regression coefficient $B_B = 4.16$; the standard error is .69. With 40 groups and a single predictor, there are 38 degrees of freedom for $MS_{residual}$, and (1, 38) df for the F test for overall regression. Note that the ratio of the regression coefficient for MOTIVATC to its standard error is much larger in the disaggregated analysis, $t(384) = 21.44$, than in the aggregated analysis, $t(38) = 6.07$. The between class coefficient B_B of 4.16 indicates that if we consider only the mean weight loss per group as a function of mean motivation in the group, the mean weight loss increases by 4.16 pounds for each one-unit increase in mean motivation in the group; each group contributes equally to this result, regardless of sample size. This value is larger than the total regression coefficient $B_T = 3.27$. The reason is that a small group with both high mean motivation and high mean weight loss or a small group with both low mean motivation and low mean weight loss can have a strong positive impact on the between class regression coefficient, since there are only 40 pairs of observations (i.e., pairs of group means) in the analysis. In fact, such groups exist in the data set. The few cases within each such extreme group have less influence when treated as individual data points in the large overall sample of $n = 386$ cases in the previous disaggregated analysis.

Disaggregated Analysis With Dummy-Coded Groups

The disaggregated analysis with $(g - 1)$ dummy codes for the g groups is given in Table 14.2.1C. For the 40 groups there are 39 dummy codes. For each dummy code, members of all the groups except one are coded zero (0); the members of the remaining group are coded one (1); the 40th group is coded zero on all 39 dummy codes and serves as the base group. Choice of the base group is arbitrary in this analysis, since the dummy codes are used together to characterize the influence of group structure on the dependent variable. The pooled within-class regression coefficient for MOTIVATC is $B_W = 3.12$ with a standard error of .14. This coefficient is the weighted average of the regression coefficient of pounds lost on motivation within each group. With 39 dummy coded predictors for group membership and the MOTIVATC predictor, there are $k = 40$ predictors in all, and again $n = 386$ for the individual women. There are $(n - k - 1) = (386 - 40 - 1) = 345$ *df* for $MS_{residual}$, and (40, 345) *df* for the F test for overall regression. This analysis controls for (partials out) the differences in mean weight loss per group when the impact of motivation on weight loss is estimated. The 39 group dummy codes account for almost 30% of the variance in pounds lost ($R^2_{groups} = .297$). MOTIVATC accounts for another 40% of the variance in pounds lost ($R^2_{groups+MOTIVATC} = .704$). The ratio of the regression coefficient for MOTIVATC to its standard error in this analysis is only very slightly smaller than in the disaggregated analysis, $t(345) = 21.77$. We are comfortable in concluding that there is an impact of motivation, once mean differences in pounds lost across groups are controlled. In this particular data set, the results of this analysis differ very little from those of the disaggregated regression analysis in Table 14.2.1A; this is hardly a necessary result. In fact, it is more usual that the standard error of the predictor in question (here MOTIVATC) is larger when group structure (clustering) is taken into account than when it is not.

This last result makes it appear that clustering has little effect on the estimate of the significance of the relationship of motivation to weight loss. However, there is more to the story. These particular data were simulated to have widely varying slopes of the regression of weight loss on motivation across groups, and the analysis with the 39 dummy codes in Table 14.2.1C does not take into account these slope differences. As previously discussed, we could estimate a model in which slope differences were permitted by forming the cross-product terms between each of the dummy codes and MOTIVATC. This analysis poses complexities due to centering. We would have to change from using dummy codes to represent groups to using weighted effects codes (see Section 8.4) so that all predictors would be centered. With weighted effects codes and centered MOTIVATC, the estimate of the regression of pounds lost on MOTIVATC in the model containing all the interaction terms would be at the mean of all the groups (see West, Aiken, & Krull, 1996).

14.3 THE RANDOM COEFFICIENT REGRESSION MODEL

We now have available a newer regression model than OLS, the *random coefficient* (*RC*) *regression model*, which provides a highly flexible approach to the handling of clustered data. The RC regression model is mathematically different from OLS regression, as are the estimation procedures employed. When data are clustered, the RC model provides accurate estimates of relationships of individual level predictors to a dependent variable while at the same time taking into account clustering and providing accurate estimates of the standard errors of regression coefficients so that alpha inflation is avoided. Random coefficient regression also permits the analysis of *multilevel data* within a single regression model. Multilevel data contain predictors measured at more than one level of aggregation, for example, measures of IQ taken on individual children within a multichild family, plus a measure of total family income taken at the level of the family. The term *multilevel model* (or, equivalently, *hierarchical linear model*) is applied to these random coefficient regression applications in which there are predictors at multiple levels of aggregation (Raudenbush & Bryk, 2002). It is in this multilevel framework that the RC regression model has enjoyed extensive use in educational research, for example, in the study of the impact of type of school (public, private) on the relationship of children's socioeconomic status to mathematics achievement (Raudenbush & Bryk, 2002) and in sociology in the study of the impact of individual and contextual (e.g., societal) variables on such outcomes as contraceptive utilization by individual women (e.g., Wong & Mason, 1985). These models have an advantage, explained in Section 14.7.3, that the *cross-level interaction* between variables that occur at different levels of aggregation can be examined (for example, the interaction between family income and child IQ on educational achievement). These models are now enjoying increasing use within psychology. In this chapter we present the random coefficient model for the continuous DV. These models have also been extended to binary, ordered category, and count variables (see Snijders & Bosker, 1999, chapter 14; Wong & Mason, 1985).

14.4 RANDOM COEFFICIENT REGRESSION MODEL AND MULTILEVEL DATA STRUCTURE

A brief review of several aspects of OLS regression sets the stage for the introduction to random coefficient regression.

14.4.1 Ordinary Least Squares (Fixed Effects) Regression Revisited

Recall that in OLS regression we have a single population regression equation, which, in the one-predictor case, is as follows:

$$\underline{Y_i} = \beta_1^* X_i + \beta_0^* + \underline{\epsilon_i} \tag{14.4.1}$$

where β_0^* is the population intercept, β_1^* is the population unstandardized regression slope, and $\underline{\epsilon_i}$ is the random error in prediction for case i. (Readers should not confuse the notation β_1^* for population unstandardized regression coefficient with the notation β_1 for the sample standardized regression coefficient.)

14.4.2 Fixed and Random Variables

Following the convention of Kreft and de Leeuw (1998), we underline random variables. By *random variables* are meant variables whose values are selected at random from a probability distribution. Both the error term $\underline{\epsilon_i}$ and the dependent variable $\underline{Y}$ are random variables. Here we assume a normal probability distribution of error in the population with mean $\mu = 0$ and variance σ_ϵ^2. In OLS regression the Xs are assumed to be *fixed*, that is to take on a predetermined set of values (though in many applications of OLS regression, we do not meet this assumption). In addition, β_0^* and β_1^* are the fixed parameters of the unstandardized regression equation for the whole population regression equation. This is the source of the term *fixed effects* regression analysis applied to OLS regression.

In any single OLS regression analysis we draw a random sample of cases from the population and estimate a single regression equation for the sample:

$$\underline{Y_i} = B_1 X_i + B_0 + \underline{e_i}. \tag{14.4.2}$$

In OLS regression the sample intercept (regression constant) B_0 and slope (regression coefficient) B_1 are estimates of the fixed intercept β_0^* and slope β_1^* in the population, respectively, and are considered fixed; there is one estimate of the fixed intercept β_0^* and one estimate of the fixed slope β_1^* from the analysis; $\underline{e_i}$ is the random error associated with observation X_i in the sample.

14.4.3 Clustering and Hierarchically Structured Data

In random coefficient regression we retain the notion of a population regression equation with a population intercept and population slope. However, we add complexity in terms of the data structure, that the data are clustered into groups, or *hierarchically structured*. The clustering yields *levels* within the data structure. The lowest level of aggregation, the individual, is referred to as *level 1* or the *micro-level*. The cluster or group level is referred to as *level 2* or the *macro-level*. It is possible to have more than two levels in the data structure (e.g., children within families within neighborhoods); we limit the presentation to two levels. The clusters in any data set are assumed to be a random sample of all possible clusters in the population. For example, the 40 women's diet groups in the numerical example are assumed to be a random sample of the population of all women's diet groups. Although in OLS regression the individual case is the unit that is randomly sampled, in random coefficient regression it is the cluster or group that is randomly sampled.

14.4.4 Structure of the Random Coefficient Regression Model

The notation of random coefficient regression and multilevel modeling is somewhat arcane. In developing the RC regression model and its application to multilevel modeling, we use common notation from leading texts in the multilevel modeling field. We do so in order that the reader may refer to other sources for further information and may follow the terminology of common software for multilevel modeling. Our notation follows that of Raudenbush and Bryk (2002) and Snijders and Bosker (1999), and is very close to that of Kreft and de Leeuw (1998) and Goldstein (1995). It is summarized in Table 14.4.1.

Within a single random coefficient regression analysis, we have g groups in all. We use the subscript j to denote any one of the groups, with the particular group unspecified (i.e., group j). The group membership of each case in the analysis is identified; hence we think of case i in group j.

The RC regression model is more complex than OLS regression because the RC model addresses the group structure inherent in the data as well as both individual level and group level relationships among variables. There are three types of regression equations in the random coefficient regression model. First, there are level 1 (micro-level) regression equations, one for each group in the data set. Second, there are level 2 regression equations that carry the group structure. Third, there is an overall regression equation, the *mixed model equation*, that combines the level 1 and level 2 equations. In addition, there are a set of *variance components* that summarize the differences among the groups. In what follows we develop the RC model and its extension to

TABLE 14.4.1
Notation for Random Coefficient Regression in the Multilevel Framework

A. Coefficients in micro-level equation for group j.
$\underline{B_{0j}}$ = level 1 regression intercept in group j.
$\underline{B_{1j}}$ = level 1 regression coefficient (slope) in group j.

B. Fixed population regression coefficients: the fixed part of the model.
γ_{00} = the population regression intercept.
γ_{10} = the population regression coefficient for the regression of the dependent variable on the level 1 predictor.
γ_{01} = the population regression coefficient for the regression of the dependent variable on the level 2 predictor.
γ_{11} = the population regression coefficient for the interaction between the level 1 and level 2 variables in predicting the dependent variable.

C. Residuals and variance components: the random part of the model.
1. Residuals.
$\underline{r_{ij}}$ = level 1 error for subject i in group j (level 1 equation).
$\underline{u_{0j}}$ = random deviation of the intercept of an individual group j from the overall population intercept (level 2 equation).
$\underline{u_{1j}}$ = random deviation of the regression coefficient of an individual group j from the overall population regression coefficient (level 2 equation).

2. Variance components.
σ^2 = variance due to random error at level 1 (i.e., variance of the $\underline{r_{ij}}$).
τ_{00} = variance of the random intercepts (i.e., variance of $\underline{u_{0j}}$).
τ_{11} = variance of the random regression coefficients (i.e., variance of $\underline{u_{1j}}$).
τ_{01} = covariance between the errors of the random regression coefficients and the random regression intercepts (i.e., covariance between $\underline{u_{0j}}$ and $\underline{u_{1j}}$).

Note: Notation follows Raudenbush and Bryk (2002) and the HLM software (Raudenbush, Bryk, Cheong, & Congdon, 2001) as well as Snijders and Bosker (1999); the only difference from Kreft and de Leeuw (1998) is their use of $\underline{\epsilon_{ij}}$, rather than $\underline{r_{ij}}$, for the level 1 residual. Goldstein (1995) also uses $\underline{\epsilon_{ij}}$, rather than $\underline{r_{ij}}$; also, the τ notation for variance components is replaced with σ^2_{u0}, etc., as in the MLwiN software (Goldstein et al., 1998).

multilevel modeling with no more than one predictor at each level. We do this for ease of presentation only. Any number of predictors may be entered at each level (micro, macro) of the model.

14.4.5 Level 1 Equations

We have the following level 1 (micro-level) equations, one for each group in the analysis:

$$\begin{aligned}
&\text{level 1 equation in group 1:} && \underline{y_{i1}} = \underline{B_{11}}x_{i1} + \underline{B_{01}} + \underline{r_{i1}}.\\
&\text{level 2 equation in group 2:} && \underline{y_{i2}} = \underline{B_{12}}x_{i2} + \underline{B_{02}} + \underline{r_{i2}}.\\
&\qquad\vdots && \\
&\text{level 1 equation in group } j\text{:} && \underline{y_{ij}} = \underline{B_{1j}}x_{ij} + \underline{B_{0j}} + \underline{r_{ij}}.\\
&\text{level 2 equation in group } g\text{:} && \underline{y_{ig}} = \underline{B_{1g}}x_{ig} + \underline{B_{0g}} + \underline{r_{ig}}.
\end{aligned} \tag{14.4.3}$$

For each group j, we have made one equation specific to that group by appending the subscript j to every term in the equation: to the dependent variable $\underline{y_{ij}}$ for case i in group j, to the predictor

x_{ij} for case i in group j, to the level 1 residual or random error $\underline{r_{ij}}$ for case i in group j. The level 1 regression intercept $\underline{B_{0j}}$ and the level 1 slope $\underline{B_{1j}}$ also carry the subscript j, again showing the specificity to group j. Since we assume that the groups are a random sample from a population of groups, the intercepts and slopes of the level 1 equations for the various groups become random variables in the random coefficient regression model. Hence they are underlined in these level 1 equations. Put another way, in any single random coefficient regression analysis, we conceptually have a whole series of regression analyses, one for each group, each with its own intercept and slope. Within the single random coefficient regression analysis there is a distribution of these intercepts and a distribution of the slopes. The term *random coefficient regression* stems from the assumption that the intercept $\underline{B_{0j}}$ and the slope $\underline{B_{1j}}$ are themselves random variables.

14.4.6 Level 2 Equations

In random coefficient regression analysis, we still retain the notion that there is an overall population regression equation. The level 2 or macro-level equations express how the set of level 1 intercepts for each cluster ($\underline{B_{0j}}$) and the level 1 slopes ($\underline{B_{1j}}$) relate to the intercept and slope from the overall population regression equation. Again, we use common notation from multilevel modeling with γ_{00} (gamma zero zero) for the population regression intercept and γ_{10} for the population regression slope. These are the *fixed parameters* of the population regression equation. We assume that the level 1 (micro-level) intercepts from the various individual groups, $\underline{B_{01}}, \underline{B_{02}}, \ldots, \underline{B_{0j}}, \underline{B_{0g}}$, vary randomly in value around the population intercept γ_{00}.

We specify a level 2 model for how the intercept $\underline{B_{0j}}$ in each group relates to the population intercept γ_{00}. The equation indicates that any $\underline{B_{0j}}$ is comprised of a fixed part γ_{00} and a random part $\underline{u_{0j}}$:

(14.4.4) level 2 model, regression intercept: $\underline{B_{0j}} = \gamma_{00} + \underline{u_{0j}}$.

Once again, Eq. (14.4.4) indicates that the regression intercept $\underline{B_{0j}}$ in any particular group j is a function of the fixed population intercept γ_{00} plus a random deviation $\underline{u_{0j}} = \underline{B_{0j}} - \gamma_{00}$ of the group intercept from the overall fixed population intercept.

Similarly, we assume that the level 1 regression slopes from the various groups $\underline{B_{11}}, \underline{B_{12}}, \ldots, \underline{B_{1j}}, \underline{B_{1g}}$, vary randomly in value around the fixed population regression coefficient γ_{10}, with random deviation $\underline{u_{1j}}$ of $\underline{B_{1j}}$ from γ_{10}, yielding the following level 2 model for the regression slope:

(14.4.5) level 2 model, regression slope: $\underline{B_{1j}} = \gamma_{10} + \underline{u_{1j}}$.

The level 2 equations characterize the group structure inherent in the data, as noted in the j subscript for each group. The identity of the groups within the analysis is embodied in the level 2 equations. The clustered nature of the data is captured at level 2. Recall that in OLS regression the group structure is identified with a set of $(g - 1)$ dummy variables for g groups. The level 2 model characterized by only Eq. (14.4.4) in random coefficient regression *replaces* the $(g - 1)$ dummy codes in OLS regression (see Section 14.4.9).

There is a relationship between the random coefficient regression model and the fixed OLS regression model, further explained later. If there is no variation among the intercepts across the groups and no variation among the slopes across the groups, then the random coefficient regression model is equivalent to fixed OLS regression.

14.4.7 Mixed Model Equation for Random Coefficient Regression

The presentation of the RC model thus far makes it appear that the level 1 and level 2 equations are treated separately. In fact, they are combined to form a single RC regression equation, referred to as the *mixed model* because it "mixes" the two levels, in that it contains terms from both the level 1 and level 2 models (the term *mixed model*, extensively used in econometrics, is enjoying increasing use in psychology). If we substitute Eqs. (14.4.4) and (14.4.5) for the intercept and slope, respectively, in group j into the level 1 equation for that group j (Eq. 14.4.3) we obtain the *mixed model* form of the random coefficient model:

$$\underline{y_{ij}} = (\gamma_{00} + \underline{u_{0j}}) + (\gamma_{10} + \underline{u_{1j}})\, x_{ij} + \underline{r_{ij}};$$

$$\underline{y_{ij}} = \gamma_{00} + \underline{u_{0j}} + \gamma_{10}\, x_{ij} + \underline{u_{1j}}\, x_{ij} + \underline{r_{ij}};$$

$$\underline{y_{ij}} = \gamma_{10}\, x_{ij} + \gamma_{00} + (\underline{u_{0j}} + \underline{u_{1j}}\, x_{ij} + \underline{r_{ij}}). \tag{14.4.6}$$

The final expression (Eq. 14.4.6) is the *mixed model equation.* It gives the regression of the Y on the level 1 predictor in terms of fixed population values, the population intercept γ_{00} and population slope γ_{10}, plus a complex error term that includes the level 1 error $\underline{r_{ij}}$ plus the level 2 deviations $\underline{u_{0j}}$ and $\underline{u_{1j}}$. The residual $\underline{r_{ij}}$ has the same interpretation as the residual in OLS regression, the extent to which the DV is not predicted from the level 1 predictor(s). The $\underline{u_{0j}}$ deviation measures the discrepancy between the specific group intercept and the fixed population intercept. The $\underline{u_{1j}}$ deviation measures the discrepancy between the specific group slope and the fixed population slope. The mixed model error term is larger than the error term for a corresponding disaggregated OLS total regression equation that ignores group membership and predicts $\underline{y_{ij}}$ from x_{ij}; the OLS error term in Eq. (14.4.1) only consists of $\underline{\epsilon_i}$, which is the analog of $\underline{r_{ij}}$ in the RC model. Equation (14.4.6) characterizes the outcome of RC regression analysis, a single regression equation that takes into account group structure in the estimation of the regression coefficient for the regression of the criterion on the level 1 predictor.

In the example of predicting pounds lost from motivation, Eq. (14.4.6) gives the regression of weight loss on motivation, taking into account the group structure of the data. The regression coefficient and intercept resulting from a single predictor RC regression as in Eq. (14.4.6) may be highly similar to the coefficient and intercept from a total regression disaggregated OLS regression analysis of the whole data set, ignoring group membership. However, the standard error of the intercept $\hat{\gamma}_{00}$ from the RC regression equation is expected to be larger (appropriately so) than in the OLS regression equation, so long as the groups exhibit variation in their individual intercepts, $\underline{B_{01}}, \underline{B_{02}}, \ldots, \underline{B_{0j}}, \underline{B_{0g}}$. Similarly, the standard error of the slope coefficient $\hat{\gamma}_{10}$ from the multilevel formulation is expected to be larger (appropriately so) than in the OLS regression equation, so long as the groups exhibit variation in their individual slopes $\underline{B_{11}}, \underline{B_{12}}, \ldots, \underline{B_{1j}}, \underline{B_{1g}}$.

14.4.8 Variance Components—New Parameters in the Multilevel Model

The RC regression model employs a concept not employed in OLS regression, that of *variance components.* The variance components are a hallmark of random coefficient models. In the random coefficient regression model we have three different sources of random errors or deviations : (1) the level 1 random errors, $\underline{r_{ij}}$, from random variation in the Y scores in Eq. (14.4.3); (2) the level 2 deviations of the random intercepts around the population intercept, $\underline{u_{0j}}$ in Eq. (14.4.4), and (3) the level 2 deviations of the random slopes around the population slope, $\underline{u_{1j}}$ in

Eq. (14.4.5). Each of these sources of random deviation can be summarized as a variance. First we have the level 1 variance σ^2 of the r_{ij}s. The level 1 variance is typically assumed to be constant across groups and thus bears no group subscript here. At level 2, we have the variance of the level 1 random intercepts around the population intercept, (i.e., the variance of the u_{0j}s), noted τ_{00} (tau zero zero). At level 2, we also have the variance of the level 1 random slopes around the population slopes (i.e., the variance of the u_{1j}s) noted τ_{11}. Each of these variances—that is, σ^2, τ_{00}, and τ_{11}—is a *variance component* of the random coefficient regression model.

The variance components τ_{00}, and τ_{11} provide a simple way to capture the impact of group structure on the relationship of predictors to the dependent variable. Consider τ_{00}, the variance of the level 1 random intercepts in the random coefficient regression model; to the extent that individual groups have different random intercepts (B_{0j}), the value of τ_{00} will be large. The parameter τ_{00} is an elegant construction: instead of keeping track of the individual intercepts $B_{01}, B_{02}, \ldots, B_{0j}, B_{0g}$ of the g groups, the intercepts of the g groups are replaced in the random coefficient model with a single variance component, the *variance of the intercepts*, τ_{00}. What about the variance of the slopes, τ_{11}? Instead of keeping track of the individual slopes $B_{11}, B_{12}, \ldots, B_{1j}, B_{1g}$ of the g groups, the slopes of the g groups are replaced in the random coefficient model with a single *variance of the slopes*, τ_{11}.

We can make a conceptual link between the variance components and the dummy codes for group membership in the third OLS regression analysis, described in Section 14.2.1 and illustrated in Table 14.2.1C. In the OLS framework, we had a much more cumbersome way to model the intercepts of the individual groups: We used $(g-1)$ dummy codes in the regression equation. If the groups had different means on the dependent variable (in our example, if the different groups had different average amounts of weight loss) then the set of $(g-1)$ dummy codes would account for significant variance in the dependent variable, because the set of dummy codes capture differences among the means of the groups on the dependent variable (equivalently, differences in intercepts if we think of estimating a regression equation in each group). Conceptually, we can replace the $(g-1)$ dummy codes of OLS regression with the single variance component τ_{00} in random coefficient regression. The random coefficient regression model requires only two variance components, τ_{00} and τ_{11}, to summarize all the between-group differences in intercepts and slopes, respectively. To summarize this same variance in OLS regression, we would need $(g-1)$ dummy codes to capture intercept differences, plus another $(g-1)$ interaction terms of each dummy code with the level 1 predictor to capture the variance of the slopes. Thus two terms in the RC regression model replace $2(g-1)$ terms in OLS regression to fully characterize the differences in regression equations across the groups.

As a final note, there is actually one more variance component in the RC regression model. The random slopes and intercepts may also covary. Thus a third variance component is estimated: the *covariance* between the level 1 slopes and intercepts across groups, noted τ_{01}. This term provides interesting information from a theoretical perspective. It may be that the intercept and slope are positively related; in the weight example, this would mean that the groups that showed the highest average weight loss also exhibited the strongest relationship between motivation and weight loss.

14.4.9 Variance Components and Random Coefficient versus Ordinary Least Squares (Fixed Effects) Regression

When does the RC regression model simplify to the fixed OLS regression model? In fixed effects regression analysis, we have only one variance component, σ^2, which is estimated by MS_{residual}. In RC regression, we have the level 1 variance component σ^2, plus two level 2 variance components, τ_{00} and τ_{11}. If τ_{00} and τ_{11} are equal to zero in a population, then the random coefficient regression model simplifies to fixed OLS regression. Under what circumstance would both τ_{00} and τ_{11} be zero? If the intercepts of all the level 1 regression equations in all the groups were identical, then τ_{00} would equal zero (no variance among the intercepts). If the slopes of all the level 1 regression equations in all the groups were identical, then τ_{11} would equal zero (no variance among the slopes). If both τ_{00} and τ_{11} are zero, then there is no effect of clustering or group membership on the outcome of the regression equation; the random coefficient regression equation is equivalent to an OLS regression that ignores group membership.

The Random Intercept Model

It is possible that either τ_{00} or τ_{11} is equal to zero but the other variance component is not. For example, suppose we estimate an RC regression model with a single level 1 predictor and with both *random intercepts and slopes*, as in Eqs. (14.4.3), (14.4.4), (14.4.5), and find that τ_{00} is greater than zero but τ_{11} equals zero. (Statistical inference for variance components is illustrated and explained later.) Again, that τ_{00} is greater than zero indicates that the intercepts differ across groups. That τ_{11} equals zero signifies that the slopes across groups are all equal to one another and can be considered as fixed (i.e., to take on a constant value across groups). Figure 14.2.1(B) illustrates this data structure. In the weight loss example, we would say that the groups differed in pounds lost (intercept) but reflected a constant relationship of MOTIVATC to pounds lost (slope). We could respecify the model as a *random intercept* model, with Eqs. (14.4.3) and (14.4.4) as before, but Eq. (14.4.5) replaced by the expression $B_{1j} = \gamma_{10}$. The random intercept model is the analog in RC regression to the OLS (fixed effects) regression analysis in which $(g - 1)$ dummy codes represent differences among the means of g groups. We consider the choice between the random intercept and fixed model in Section 14.14.

14.4.10 Parameters of the Random Coefficient Regression Model: Fixed and Random Effects

We can organize all the parameters of the RC regression model into two classes, referred to as the *fixed effects* and the *random effects*. The population regression intercept and slope (regression coefficient), γ_{00}, γ_{10}, respectively, in Eqs. (14.4.4), (14.4.5), and (14.4.6) are referred to as the *fixed effects*. The variance components are referred to as the *random effects*. These two classes of parameters are two distinct foci of hypothesis testing in the random coefficient regression model, as we illustrate later.

14.5 NUMERICAL EXAMPLE: ANALYSIS OF CLUSTERED DATA WITH RANDOM COEFFICIENT REGRESSION

CH14EX03

We now use what we have learned about random coefficient regression. We return to the weight reduction example introduced in Section 14.2.1 and employ RC regression to predict pounds lost from motivation. Random coefficient models were executed in SAS PROC MIXED; Singer (1998) has provided an excellent tutorial on the use of PROC MIXED for analyzing multilevel models and longitudinal growth models. Results are presented in Table 14.5.1, which shows output from SAS PROC MIXED.

TABLE 14.5.1
Analyses of Weight Loss as a Function of Motivation With Random Coefficient Regression

A. Random coefficient regression: unconditional cell means model to derive intraclass correlation.

Random Part: Covariance Parameter Estimates (REML)

	Cov Parm	Subject	Estimate	Std Error	Z	Pr > $\|Z\|$
$\hat{\tau}_{00}$, variance of intercepts	UN(1, 1)	GROUP	4.906	1.560	3.14	0.002
$\hat{\sigma}^2$, level 1 residual	Residual		16.069	1.225	13.12	0.001

B. Random coefficient regression: prediction of pounds lost from motivation (level 1) at outset of diet program

Random Part: Covariance Parameter Estimates (REML)

	Cov Parm	Subject	Estimate	Std Error	Z	Pr > $\|Z\|$
$\hat{\tau}_{00}$, variance of intercepts	UN(1, 1)	GROUP	2.397	0.741	3.23	0.001
$\hat{\tau}_{01}$, covariance between slope and intercept	UN(2, 1)	GROUP	0.585	0.385	1.52	0.128
$\hat{\tau}_{11}$, variance of slopes	UN(2, 2)	GROUP	0.933	0.376	2.48	0.013
$\hat{\sigma}^2$ level 1 residual	Residual		5.933	0.476	12.47	0.001

Solution for Fixed Effects

	Effect	Estimate	Std Error	DF	t	Pr > $\|t\|$
$\hat{\gamma}_{00}$	NTERCEPT	15.138	0.280	39	54.13	0.001
$\hat{\gamma}_{10}$	MOTIVATC	3.118	0.211	345	14.80	0.001

Note: The UN notation is the SAS notation for a variance component. REML stands for restricted maximum likelihood (Section 14.10.1).

14.5.1 Unconditional Cell Means Model and the Intraclass Correlation

We begin with estimation of the ICC in RC regression as an alternative to the approach shown for estimating the ICC in the fixed model (Section 14.1.2). Actually, the RC model used for estimating the ICC is a bit simpler than the one presented in Eqs. (14.4.3), (14.4.4), and (14.4.5). It contains no level 1 predictor X in Eq. (14.4.3), so the level 1 equation corresponding to Eq. (14.4.3) becomes $y_{ij} = B_{0j} + r_{ij}$ for group j. Each group's equation has an intercept B_{0j} but no slope; each individual score is predicted solely from the mean of the group. The hierarchical structure of identifying group membership is retained in the level 2 equation for the intercept, which is identical to Eq. (14.4.4). Since there is no longer a predictor X and thus no longer a random slope B_{1j} in the model, there is no equation Eq. (14.4.5). This model is called an *unconditional cell means model*. It is equivalent to a one-factor *random effects* ANOVA of pounds lost with group as the sole factor;[4] the 40 groups become the 40 levels of the group

[4]The analysis of variance (ANOVA) model in common use is the *fixed effects ANOVA* model. Consider a one-factor model. A fixed effects one-factor ANOVA assumes that the levels of the factor included in the data set are a fixed set of levels, in fact, all the levels to which we wish to generalize. This corresponds to the fixed effects OLS regression

factor. This analysis provides estimates of $\hat{\tau}_{00}$, the variance among the 40 group intercepts (or equivalently, the variance among the groups in mean number of pounds lost), and $\hat{\sigma}^2$, the level 1 residual variance. (The $\hat{\tau}_{00}$ here is the same parameter as τ in Section 14.1.2; that is, $\hat{\tau}_{00}$ and τ measure the same thing. What differs here is the estimation approach.) As shown in Table 14.5.1A, the estimates of $\hat{\tau}_{00}$ and $\hat{\sigma}^2$ are 4.906 and 16.069, respectively. They reflect 4.906 units of between class variance in pounds lost (differences between groups in mean pounds lost by the end of the experiment) and 16.069 units of within class variance in pounds lost that might be accounted for by the treatment and motivation predictors. Both these values are significantly greater than zero, according to the z tests reported in Table 14.5.1A (i.e., $z = 3.14, z = 13.12$, for $\hat{\tau}_{00}$ and $\hat{\sigma}^2$, respectively). That the value of $\hat{\tau}_{00}$ is significantly greater than zero tells us that there is random variation among the intercepts of the individual groups—we should not ignore clustering. We will use the values of $\hat{\tau}_{00}$ and $\hat{\sigma}^2$ to track the impact of the level 1 and level 2 predictors in accounting for weight loss. These two values are used to compute the ICC, which is given as $\text{ICC} = \hat{\tau}_{00}/(\hat{\tau}_{00} + \hat{\sigma}^2) = 4.906/(4.906 + 16.069) = .24$ (this is the same as the ICC formula in Eq. 14.1.1). The estimated ICC of .24 is very substantial. This estimate is very similar to that derived from the fixed effects ANOVA in Section 14.2.1 (ICC estimate of .22). We expect that if disaggregated OLS regression and RC regression are applied to these data, the standard errors of regression coefficients in OLS will be smaller than in RC regression, leading to overestimates of significance of predictors.

14.5.2 Testing the Fixed and Random Parts of the Random Coefficient Regression Model

In a next step we examine the prediction of pounds lost from motivation in an RC regression model, given in Table 14.5.1B. There are two parts to the analysis: (1) a *random part* that provides the estimates of the variance components, $\hat{\tau}_{00}, \hat{\tau}_{01}, \hat{\tau}_{11}$, and $\hat{\sigma}^2$, and (2) a *fixed part* that provides the estimates $\hat{\gamma}_{00}$ and $\hat{\gamma}_{10}$ of the fixed regression constant and fixed regression slope, γ_{00} and γ_{10}, respectively. We consider first the fixed part, given under Solution for Fixed Effects in Table 14.5.1B. The RC regression equation predicting pounds lost from motivation is $\hat{Y} = 3.12\,\text{MOTIVATC} + 15.14$, where $\hat{\gamma}_{00} = 15.14$, and $\hat{\gamma}_{10} = 3.12$. For every 1-unit increase in motivation (on a 6-point scale), predicted pounds lost increases by 3.12 pounds, with an average weight loss per group of 15.14 pounds. Note the standard error of the random regression coefficient is .211.

Now we consider the random part. In the model in Table 14.5.1A, which contained no level 1 predictor, there are only two variance components, τ_{00} and σ^2. When motivation is added as a level 1 predictor, the possibility arises that the slope of the regression of pounds lost on motivation may vary across groups; hence the model presented in Table 14.5.1B contains two more variance components, the variance of the slopes across groups, τ_{11}, and the covariance

model, which assumes that predictor X takes on a fixed set of values, and that our results pertain only to those values. The *random effects ANOVA model* assumes that the levels of the factor included in the data set are a random sample of all possible levels to which we wish to generalize. Consider the 40 women's groups in the diet example. If we create a fixed factor of Women's Group in a fixed effects ANOVA, with the 40 women's groups as 40 levels of the factor (as we did to estimate the ICC from the one-factor fixed effects ANOVA in Section 14.2.1), we should conceptualize any results as pertaining to only those 40 groups. If we create a random factor of Women's Group in a random effects ANOVA, with the 40 women's groups as a random sample of possible levels of the factor, then we may generalize our results to the "population of women's groups." Historically, in the 1960s and 1970s, much attention was paid to the distinction between fixed and random ANOVA models. This distinction has not been a focus for over 20 years; researchers in psychology, at least, have automatically used fixed effects ANOVA, perhaps without awareness of the distinction from random effects ANOVA (software packages for ANOVA as a default provide fixed effects ANOVA). Now that random coefficient regression models are becoming more popular, there is a new awareness of the distinction between fixed and random models.

between the intercept and slope, τ_{01}. The estimate of the variance of the slopes across groups is $\hat{\tau}_{11} = .933$, and is significantly greater than zero, $z = 2.48$; this tells us that the groups differ in the slopes of pounds lost on motivation, and that once again, the clustering should not be ignored. The covariance between the random slopes and random intercepts of the 40 groups τ_{01} is estimated as $\hat{\tau}_{01} = .585$. The positive covariance tells us that the higher the intercept for a group (the higher the pounds lost), the higher the slope, or the stronger the positive relationship of pounds lost to motivation. However, this covariance does not differ significantly from zero, $z = 1.52$, ns.

What has happened to the estimated within group variance in pounds lost $\hat{\sigma}^2$ with the addition of the level 1 predictor motivation? Recall that $\hat{\sigma}^2 = 16.069$ in the unconditional means model without the motivation predictor in Table 14.5.1A. The addition of motivation as a predictor yields a substantially reduced $\hat{\sigma}^2 = 5.933$ (Table 14.5.1B); in all $[(16.069 - 5.933)/16.069] \times 100 = 63\%$ of the within group variance in pounds lost has been accounted for by the level 1 motivation variable. Similarly, consider $\hat{\tau}_{00} = 4.906$ in the unconditional means model without the motivation predictor (Table 14.5.1A). The addition of motivation as a predictor yields a substantially reduced $\hat{\tau}_{00} = 2.397$ (Table 14.5.1B); in all $[(4.906 - 2.397)/4.906] \times 100 = 51\%$ of the between-group differences in average number of pounds lost has been accounted for by the level 1 motivation variable. The 40 groups fluctuated in mean number of pounds lost; this fluctuation was systematically related to the mean motivation level of individuals within the groups.

14.6 CLUSTERING AS A MEANINGFUL ASPECT OF THE DATA

To this point we have treated clustering or group structure in the data as if it were a nuisance that posed problems when we study the relationship of a level 1 predictor to a dependent variable (here, MOTIVATC to pounds lost). Clustering in data arises for meaningful reasons occasioned by the nature of research questions. For example, what is the effect of a mother's interest in literature on her children's development of reading habits? Or, is the relationship between students' mathematics background and their performance in graduate courses in applied statistics affected by the characteristics of the instructor?

In our study of weight loss, we might consider how level of group cohesiveness (level 2) and women's motivation (level 1) affect the number of pounds lost, a level 1 outcome. We may further ask whether there is a *cross-level interaction* between group cohesion and individual women's motivation in predicting individual women's pounds lost; perhaps cohesiveness in the group enhances the relationship of motivation to weight loss among individual women. Our research questions reflect this multilevel structure.

14.7 MULTILEVEL MODELING WITH A PREDICTOR AT LEVEL 2

To translate the preceding examples into the RC regression model we add predictors at level 2. The term *multilevel modeling* is typically applied here, when we have both level 1 and level 2 predictors. In the weight loss example, we might add a measure of group cohesiveness W as a level 2 predictor.

14.7.1 Level 1 Equations

When we add a level 2 predictor to the regression equations the level 1 equation is unaffected, remaining the same as earlier:

(14.4.3) level 1 equation in group j: $\underline{y_{ij}} = \underline{B_{1j}} x_{ij} + \underline{B_{0j}} + \underline{r_{ij}}$.

14.7.2 Revised Level 2 Equations

The level 2 equations show the addition of the level 2 predictor W_j. This predictor has one score for each group (e.g., a measure of group cohesion). If we believe that the intercepts of the groups are affected by the level 2 predictor, we add the level 2 predictor to Eq. (14.4.4). For example, if we believe that the average pounds lost per group depended on group cohesion (a source of mutual social support for dieting), we would add the group cohesion predictor W_j to the level 2 equation for the random intercept. Note that there is only one score on group cohesion for each group; that is, cohesion is a characteristic of the group. We might hypothesize that higher group cohesion is associated with greater average weight loss in a group. Average weight loss per group is reflected in the intercept for the group $\underline{B_{0j}}$.

(14.7.1) level 2 equation, intercept: $$\underline{B_{0j}} = \gamma_{01} W_j + \gamma_{00} + \underline{u_{0j}}.$$

Note that we have a new fixed effect, γ_{01}, the regression coefficient of the group intercept $\underline{B_{0j}}$ on the level 2 predictor.

If we believed that the relationship of the dependent variable Y to the level 1 predictor is affected by the level 2 predictor, we add the level 2 predictor to Eq. (14.4.5). For example, we might hypothesize that high group cohesion would strengthen the relationship of motivation to weight loss, which is reflected in the level 1 regression coefficient $\underline{B_{1j}}$ for each group. We would then add the group cohesion measure W_j to the level 2 slope equation:

(14.7.2) level 2 equation, slope: $$\underline{B_{1j}} = \gamma_{11} W_j + \gamma_{10} + \underline{u_{1j}}.$$

Again, we have a new fixed effect, γ_{11}, the regression coefficient of the group random slope $\underline{B_{1j}}$ on the level 2 predictor.

14.7.3 Mixed Model Equation With Level 1 Predictor and Level 2 Predictor of Intercept and Slope and the Cross-Level Interaction

Once again, we write a mixed model equation that shows the prediction of Y from the level 1 and level 2 predictors. We substitute the level 2 equations (Eqs. 14.7.1 and 14.7.2) into the level 1 equation (Eq. 14.4.3) to obtain a mixed model equation:

$$\underline{y_{ij}} = (\gamma_{01} W_j + \gamma_{00} + \underline{u_{0j}}) + (\gamma_{11} W_j + \gamma_{10} + \underline{u_{1j}})\, x_{ij} + \underline{r_{ij}}$$

$$\underline{y_{ij}} = \gamma_{01} W_j + \gamma_{00} + \underline{u_{0j}} + \gamma_{11} W_j x_{ij} + \gamma_{10} x_{ij} + \underline{u_{1j}} x_{ij} + \underline{r_{ij}}$$

(14.7.3) $$\underline{y_{ij}} = \gamma_{01} W_j + \gamma_{10} x_{ij} + \gamma_{11} W_j x_{ij} + \gamma_{00} + (\underline{u_{0j}} + \underline{u_{1j}} x_{ij} + \underline{r_{ij}}).$$

Cross-Level Interaction

The first three terms of Eq. (14.7.3) indicate that the pounds lost criterion is predicted from the level 2 cohesion predictor W_j, the level 1 centered motivation predictor x_{ij}, and the *cross-level interaction* $W_j x_{ij}$ between cohesion W_j and centered motivation x_{ij}. The cross-level interaction in the mixed model expression for the slope (Eq. 14.7.3) results from the level 2 equation for the slope (Eq. 14.7.2), which states that the level 2 predictor predicts the level 1 slope. In causal terms, the level 2 predictor changes (or moderates) the relationship of the level 1 predictor to the dependent variable (see Section 7.3.2 for a discussion of moderation). This *fixed part* of the equation is analogous to the OLS regression equation with a two-predictor interaction in Eq. (7.1.2); there are now four *fixed effects*—$\gamma_{01}, \gamma_{10}, \gamma_{11}, \gamma_{00}$—in the multilevel model. They correspond to the regression coefficients and regression intercept, respectively, in OLS regression equation $\hat{Y} = B_1W + B_2X + B_3WX + B_0$, containing a WX interaction.

Variance Components

The final term in Eq. (14.7.3) contains the random components of the model. It is once again a complex error term, characteristic of RC regression equations. It contains the same components as in the model with no level 2 predictor, in Eq. (14.4.6), the level 1 residual r_{ij} plus level 2 residuals u_{0j} and u_{1j} for the intercept and slope, respectively. The level 1 residual r_{ij} retains the same interpretation as in Eq. (14.4.6). However, the interpretation of the level 2 random error terms u_{0j} and u_{1j} changes when a level 2 predictor is added. The term u_{0j} is now the residual deviation of the level 1 intercept B_{0j} from the population intercept γ_{00} after the level 1 intercept has been predicted from the level 2 predictor. Put another way, u_{0j} measures the part of the discrepancy between the group j intercept B_{0j} and the population intercept γ_{00} that cannot be predicted from the level 2 predictor. If the level 2 predictor W provides prediction of the level 1 intercept, then u_{0j} in Eq. (14.7.1) will be smaller than in Eq. (14.4.4) in a model with no level 2 predictor. In the weight loss example, with group cohesion as the level 2 predictor, u_{0j} represents the part of the level 1 intercept that cannot be accounted for by group cohesion. The same is true for residual u_{1j} in Eq. (14.7.2). If the level 2 predictor explains some of the variance in the level 1 slopes, then the residual in Eq. (14.7.2) will be smaller than in Eq. (14.4.5).

Finally, instead of considering the residuals u_{0j} and u_{1j}, we can think of their variances, τ_{00}, and τ_{11}, the variance components in the model. *A goal of multilevel modeling is to account for these variances of the random intercepts and slopes, respectively, by level 2 predictors (i.e., to explain the differences among the groups in their intercepts and slopes).*

14.8 AN EXPERIMENTAL DESIGN AS A MULTILEVEL DATA STRUCTURE: COMBINING EXPERIMENTAL MANIPULATION WITH INDIVIDUAL DIFFERENCES

We have strong interest in psychology in the effects of experimental manipulations. Moreover, we have interest in how individual differences interact with experimental manipulations. We often examine whether individuals respond to an experimental manipulation or intervention more strongly or weakly as a function of some stable individual difference characteristics. In multilevel analysis the experimental manipulation and individual difference characteristics are readily portrayed, with participation in conditions of an experiment treated as a level 2 variable and individual characteristics that may affect how individuals respond to the conditions as level 1 variables. Note that if there is no clustering in the data set, then a fixed OLS regression approach can be taken instead. The experimental manipulation (e.g., treatment versus control), the individual difference variable, and the manipulation by individual difference interaction serve as predictors. Note that this is not the classic ANCOVA model described in Section 8.7.5, which assumes no interaction between the manipulation and individual difference covariate. Rather it is the aptitude-treatment or experimental personality design considered in West, Aiken, and Krull (1996; see also Section 9.3).

Again consider the diet example. Now assume that the 40 women's groups are entered into a research project to evaluate the impact of a diet program (the "treatment") on the number of pounds lost in a three-month period. Groups are randomly assigned to experimental condition (i.e., the level 2 unit is the unit of random assignment). There are two conditions: (1) baseline condition (control group) of weekly group meetings to discuss dieting triumphs and tragedies, or (2) a multicomponent treatment (experimental group) consisting of diet specification, weigh-in and counseling, exercise, food preparation lessons, plus weekly group meetings to discuss dieting triumphs and tragedies. As before, at the individual level, the motivation of each individual to lose weight at the outset of the intervention is assessed as the level 1 predictor of weight loss. The treatment condition (experimental versus control) is a level 2 predictor called TREATC. Issues of appropriate coding of the treatment variable and centering the motivation variable are considered in Section 14.9.

The mixed model regression equation for the experiment is as follows:

$$\underline{y_{ij}} = \gamma_{01}\ \text{TREATC}_j + \gamma_{10}\ \text{MOTIVATC}_{ij} + \gamma_{11}\ \text{TREATC}_j \times \text{MOTIVATC}_{ij} + \gamma_{00} + (\underline{u_{0j}} + \underline{u_{1j}}\, x_{ij} + \underline{r_{ij}}) \tag{14.8.1}$$

In a multilevel analysis we would learn whether the treatment (TREATC) had an effect on weight loss (experimental manipulation), whether motivation (MOTIVATC) predicted weight loss (individual difference), and whether there was a cross-level interaction between treatment and motivation (TREATC × MOTIVATC). We might hypothesize that treatment strengthens the relationship of motivation to pounds lost by giving motivated participants the vehicles for effective dieting. In terms of level 2 Eq. (14.7.2), we are hypothesizing that treatment moderates the relationship of motivation to weight loss, an interaction hypothesis. This example provides a model for analysis of the many experimental studies in which experimental conditions are administered to groups of subjects and in which either group composition or group processes (e.g. increased cohesion), could affect outcomes.

The mixed model framework can accommodate a second interpretation of the cross-level interaction as well. We can conceptualize that the level of motivation moderates the impact of treatment. This conceptualization is consistent with research that asks, "for whom is treatment most effective," under the assumption that characteristics of the individual (e.g., motivation) condition the impact of treatment. This second conception does not fit into the hierarchical structure that the level 2 variable affects the level 1 variable. Nevertheless, the cross-level interaction in the mixed model framework can accommodate this interpretation.

14.9 NUMERICAL EXAMPLE: MULTILEVEL ANALYSIS

We now explore the analysis of the diet treatment, assuming that 60% of the 40 groups, or 24 groups (comprised of $n = 230$ individual cases in all) were randomly assigned to the treatment, the other 16 groups (comprised of $n = 156$ individual cases in all), to control. Treatment is centered around the grand mean at the individual level into a weighted effects coded predictor[5] TREATC, where experimental $= 156/(156 + 230) = .404$, and

[5] Coding and centering (scaling) issues become very complex in multilevel models with cross level interactions in which there are unequal sample sizes in each group. Since the regression model contains an interaction, the first order effects of MOTIVATC and TREATC are conditional (i.e., are interpreted at the value of zero on the other predictor), just as in OLS regression with interactions (see Section 7.12). The scaling of the IVs to produce appropriate 0-points on each variable will depend on the sampling plan and the effect(s) of most interest in the study. In the present case, we centered MOTIVATC around the grand mean of all of the individual cases. A weighted effects code based on the number of cases in the treatment and control at the individual level was used to create TREATC. This choice parallels the centering of MOTIVATC and thus facilitates interpretation. The interpretation of each conditional main effect is at the mean of all *individual* cases on the other predictor. This choice also permits the most direct comparison of the present multilevel results with those of the original OLS analysis which ignores group membership.

TABLE 14.9.1
Multilevel Analysis and OLS Regression Analysis of the Weight Experiment With an Intervention and an Individual Difference Variable

A. Multilevel random coefficient regression: prediction of pounds lost from motivation (level 1), treatment (level 2) and the cross-level interaction between motivation and treatment

	Random Part: Covariance Parameter Estimates (REML)					
	Cov Parm	Subject	Estimate	Std Error	Z	Pr > $\|Z\|$
$\hat{\tau}_{00}$, variance of intercepts	UN(1, 1)	GROUP	1.967	0.657	2.99	0.003
$\hat{\tau}_{01}$, covariance between slope and intercept	UN(2, 1)	GROUP	0.145	0.314	0.46	0.645
$\hat{\tau}_{11}$, variance of slopes	UN(2, 2)	GROUP	0.556	0.301	1.85	0.065
$\hat{\sigma}^2$ level 1 residual	Residual		5.933	0.475	12.48	0.001

		Solution for Fixed Effects				
	Effect	Estimate	Std Error	df	t	Pr > $\|t\|$
$\hat{\gamma}_{00}$	INTERCEPT	15.166	0.259	38	58.49	0.001
$\hat{\gamma}_{01}$	TREATC	1.528	0.529	38	2.89	0.006
$\hat{\gamma}_{10}$	MOTIVATC	3.130	0.185	344	16.95	0.001
$\hat{\gamma}_{11}$	TREATC*MOTIVATC	1.245	0.377	344	3.30	0.001

B. OLS regression: prediction of pounds lost from motivation, treatment and the interaction between motivation and treatment

Effect	Estimate	Std Error	df	t	Sig t
INTERCEPT	15.105	.148			
TREATC	1.578	.301	1	5.239	.000
MOTIVATC	3.330	.145	1	22.968	.000
TREATC*MOTIVATC	1.446	.300	1	4.820	.000

Note: MOTIVATC and TREATC are the motivation and treatment predictors each centered around the grand mean of the variable. SAS reports Sig t as .000; in publication report $p < .001$.

control $= (-230)/(156 + 230) = -.596$. Table 14.9.1A provides the analysis of the multilevel (mixed) model Eq. (14.8.1). The fixed part of the analysis yields the regression equation:

$$\hat{Y} = 1.53 \text{ TREATC} + 3.13 \text{ MOTIVATC} + 1.25 \text{ TREATC} \times \text{MOTIVATC} + 15.17$$

The positive interaction of treatment with motivation indicates a synergy between individual level motivation and treatment. (See Section 9.3, for a treatment of continuous by categorical variable interactions.)

We first take the interpretation of the cross-level interaction that is consistent with the multilevel formulation, that the level 2 predictor (treatment) affects the random intercepts (average pounds lost per group), and that the level 2 predictor also modifies the random slopes (the relationship between motivation and pounds lost).

We form the simple regression equations for the regression of pounds lost on motivation within each treatment condition, just as is done in OLS regression (see Section 7.3). To do so, we rearrange the regression equation to obtain the simple slope of pounds lost on motivation as a function of treatment.

$$\hat{Y} = (3.13 + 1.25\ \text{TREATC})\ \text{MOTIVATC} + 1.53\ \text{TREATC} + 15.17.$$

For the control group, where TREATC $= -.596$, the regression of pounds lost on motivation is as follows:

$$\hat{Y} = [3.13 + 1.25(-.596)]\text{MOTIVATC} + 1.53(-.596) + 15.17$$

$$\hat{Y} = 2.38\ \text{MOTIVATC} + 14.26.$$

For the experimental group, where TREATC $= .404$, the regression of pounds lost on motivation is as follows:

$$\hat{Y} = [3.13 + 1.25(.404)]\text{MOTIVATC} + 1.53(.404) + 15.17$$

$$\hat{Y} = 3.64\ \text{MOTIVATC} + 15.79.$$

The simple regression equations are illustrated in Fig. 14.9.1(A). The treatment raised the average number of pounds lost at the mean motivation level of the 386 cases from 14.26 to 15.79 pounds. Further, it strengthened the relationship of motivation to pounds lost. In the control group, each 1 unit increase in motivation was associated with a predicted 2.38 pounds lost. In the experimental group, each 1 unit increase in motivation was associated with a predicted 3.64 pounds lost.

We now take the second interpretation of the interaction, asking whether the impact of treatment varies at different levels of motivation. We rearrange the regression equation to obtain simple regression equations of pounds lost on treatment at different levels of motivation (see also Aiken & West, 1991, Chapter 7; West, Aiken, & Krull, 1996.)

$$\hat{Y} = (1.53 + 1.25\ \text{MOTIVATC})\ \text{TREATC} + (3.13\ \text{MOTIVATC} + 15.17)$$

Motivation is centered at the grand mean of the 386 cases, so that $M_{\text{MOTIVATC}} = 0.00$; and $sd_{\text{MOTIVATC}} = 1.02$. The regression of pounds lost on treatment is shown next at one standard deviation below the mean of motivation, at the mean of motivation, and one standard deviation above the mean of motivation:

$$\text{1 } sd \text{ below:}\quad \hat{Y} = [(1.53 + (1.25)(-1.02)]\text{TREATC} + [(3.13)(-1.02) + 15.17]$$

$$\hat{Y} = .26\ \text{TREATC} + 11.97.$$

$$\text{at the mean:}\quad \hat{Y} = 1.53\ \text{TREATC} + 15.17.$$

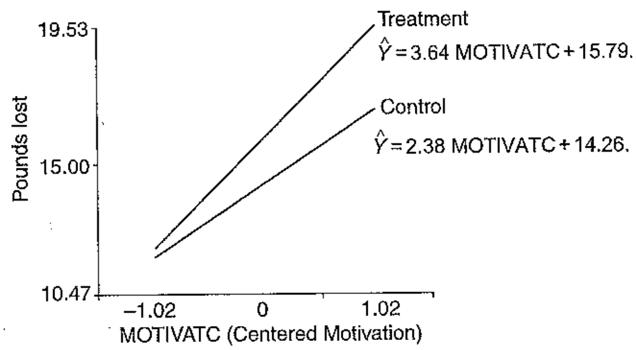
$$\text{1 } sd \text{ above:}\quad \hat{Y} = [(1.53 + (1.25)(1.02)]\text{TREATC} + [(3.13)(1.02) + 15.17]$$

$$\hat{Y} = 2.80\ \text{TREATC} + 18.36.$$

These simple regression equations are illustrated in Fig. 14.9.1(B). As individual motivation increases, the impact of treatment on weight loss increases as well. There is essentially no impact of treatment when motivation is low, with an increase to 2.80 pounds of weight loss attributable to treatment when motivation is high.

The variances of the intercepts and slopes are conceptualized as sources of variance to be accounted for by level 2 predictor TREATC. Recall that the intercept from each individual group (with motivation centered) reflects the amount of weight lost in that group; if the intervention has an effect, the variance in intercepts should at least in part be accounted for by the treatment

(A) Regression of pounds lost on motivation (centered MOTIVATC) as a function of treatment condition.



(B) Impact of treatment (centered TREATC) on pounds lost as a function of motivation level.

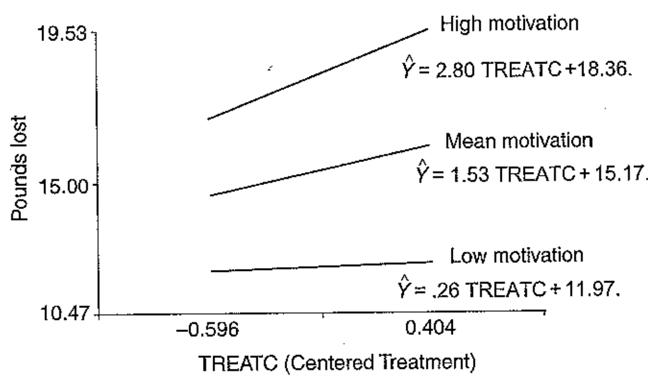


FIGURE 14.9.1 Simple slopes for cross-level interaction between MOTIVATC (level 1) and TREATC (level 2) predictors.

condition. The random part of the analysis in Table 14.9.1A versus that in Table 14.5.1B shows that all three level 2 variance components have been reduced by the addition of the level 2 treatment variable plus the cross-level interaction. The variance of the intercepts is in part accounted for by the treatment, with $\hat{\tau}_{00}$ dropping from 2.397 to 1.967, or a $[(2.397 - 1.967)/.2.397] \times 100 = 18\%$ reduction in variance unaccounted for intercept variance. There is still a significant amount of intercept variance remaining to be accounted for, $z = 2.99$. The variance of the slopes is well accounted for by treatment, with $\hat{\tau}_{11}$ dropping from .933 to .556, or a $[(.933 - .556)/.933] \times 100 = 35\%$ reduction in unaccounted for slope variance; there is, however, unaccounted for variance remaining in the slopes, $z = 1.85, p = .06$.

Table 14.9.1B gives the disaggregated OLS regression analysis ignoring group membership, for comparison with the random coefficient regression analysis in Table 14.9.1A. Pounds lost is again predicted from TREATC, MOTIVATC, and their interaction. The regression equation is almost identical in the two analyses; for the OLS analysis,

$$\hat{Y} = 1.58 \text{ TREATC} + 3.33 \text{ MOTIVATC} + 1.45 \text{ TREATC} \times \text{MOTIVATC} + 15.10.$$

However, the standard errors are uniformly smaller in OLS than in the random coefficient model, leading to alpha inflation in significance tests. Of particular note is that the standard error for the level 2 TREATC variable is .529 in the multilevel model in Table 14.9.1A and only .301 in the disaggregated OLS regression equation in Table 14.9.1B. The multilevel analysis handles the TREATC variable as if it were based on 40 observations (one per group). The OLS analysis handles the TREATC variable as if it were based on 386 independent cases, yielding a negatively biased standard error. Overall, the difference in standard errors between the single-level OLS model and the multilevel model can be attributed to two sources. First is the change in model from one that ignores clustering (single-level OLS regression) to a model that accounts for clustering (multilevel RC regression). Second is the difference in estimation procedures for OLS versus the multilevel model, described in the next section.

14.10 ESTIMATION OF THE MULTILEVEL MODEL PARAMETERS: FIXED EFFECTS, VARIANCE COMPONENTS, AND LEVEL 1 EQUATIONS

14.10.1 Fixed Effects and Variance Components

The approach to estimation is a key difference between OLS regression and RC regression. The parameter estimates in RC regression are obtained by maximum likelihood estimation (ML), described in Section 11.2 and in Section 13.2.9, or alternatively by a related method, restricted maximum likelihood (REML). Chapter 3 of Raudenbush and Bryk (2002) provides detail on estimation of the multilevel model and on hypothesis testing within the multilevel model. The fixed and random parts of the model (i.e., the fixed effects and variance components) are estimated using iterative procedures. The estimation begins with an initial estimate of one set of parameters (say, the fixed effects) and uses these values in the estimation of the other set (the variance components). The new estimates of the second set are used to update those of the first set, and the procedure continues in this manner until the process converges. (Convergence of iterative solutions is explained in Section 13.2.9.) Estimation of variance components involves algorithms that produce maximum likelihood estimates.

Confidence intervals (interval estimates) are available for the fixed effects and for the variance components. The confidence intervals (and statistical tests) for variance components are problematic because the sampling distributions of the variance components are skewed.

14.10.2 An Equation for Each Group: Empirical Bayes Estimates of Level 1 Coefficients

To this point, we have ignored the regression equations within each group. Yet a third class of parameters may be estimated—these are the level 1 *random intercept* $\underline{B_{0j}}$ and *random slope* $\underline{B_{1j}}$ for each individual group. These estimators are another important contribution of the $\overline{\text{RC}}$ regression model. We have not heretofore focused on regression equations for the individual groups. In many application of RC regression and the multilevel framework, the focus is on the overall relationships of level 1 and 2 predictors to some outcome, and whether there is evidence of variability in the intercepts and slopes across groups. There is not a focus on the regression estimates $\underline{B_{0j}}$ and $\underline{B_{1j}}$ in particular groups. However, there are instances in which the estimates of regression equations for individual groups are of importance, for example, in policy-related research in which decisions are made as to what classes of individuals might receive special treatments or interventions based on evidence of the efficacy of the treatments for those particular classes.

One obvious option for obtaining the estimates is to carry out an OLS regression in each group. If group sizes are large, this is a viable strategy. But what if we have small group sizes? It is possible to have groups so small in RC regression analysis that there are fewer cases in the group than there are level 1 predictors in the RC regression equation. The OLS regression equation cannot even be estimated! Alternatively, we may have groups large enough to estimate the OLS equation, but still so small that we have little, if any, confidence in the resulting estimates. In still other groups, larger in size, the OLS estimates may be quite reliable. An approach to estimation of regression coefficients called *empirical Bayes estimation* allows us to obtain an estimate of the regression coefficients in each group. The approach actually combines estimates of the intercept and slope for each group from two different sources.

Two Estimates of Regression Coefficients for a Single Group

Assume we have a very large sample size in group j to estimate the OLS regression slope and intercept for that group. We compute an OLS regression using only the data from group j and obtain our first set of estimates. The first set of estimates are the OLS estimates, which we will call $B_{0j,\mathrm{OLS}}$ and $B_{1j,\mathrm{OLS}}$ for the estimates of the intercept and slope, respectively, in group j.

Assume now that we have no information about group j. We would use the RC regression equation, generated from all the cases from all the groups, to provide a set of estimates of the coefficients that could be used for the individual group j. This second set of estimates for individual group j is based on the estimates from the full data set of the population fixed effects, $\hat{\gamma}_{00}, \hat{\gamma}_{01}, \hat{\gamma}_{10}$, and $\hat{\gamma}_{11}$. The estimates for group j, which we will call $B_{0j,\mathrm{POP}}$ and $B_{1j,\mathrm{POP}}$ to indicate that they are based on estimates of the population fixed effects, are

$$B_{0j,\mathrm{POP}} = \hat{\gamma}_{00} + \hat{\gamma}_{01} W_j; \qquad (14.10.1)$$

$$B_{1j,\mathrm{POP}} = \hat{\gamma}_{10} + \hat{\gamma}_{11} W_j. \qquad (14.10.2)$$

where W is the level 2 predictor and W_j is the value of this predictor in group j. Note again that the estimates of the fixed population effects $\hat{\gamma}_{00}, \hat{\gamma}_{01}, \hat{\gamma}_{10}$, and $\hat{\gamma}_{11}$ are based on all the cases in the whole analysis, regardless of group membership.

Empirical Bayes Estimators

In practice we are usually somewhere between the two extreme situations so that both sets of estimates provide useful information. Estimators of the level 1 coefficients for each group j are obtained by taking a weighted average of the two sets of estimates; these estimators are termed *shrinkage estimators* or *empirical Bayes* (*EB*) estimators. We use the notation $B_{0j,\mathrm{EB}}$ and $B_{1j,\mathrm{EB}}$ for the resulting empirical Bayes estimates of the intercept and slope. Whether the estimates from the single group j, $B_{0j,\mathrm{OLS}}$ and $\hat{B}_{1j,\mathrm{OLS}}$, or the estimates from the full data set, $B_{0j,\mathrm{POP}}$ and $B_{1j,\mathrm{POP}}$, are more heavily weighted in forming $B_{0j,\mathrm{EB}}$ and $B_{1j,\mathrm{EB}}$ depends on the precision of the estimates from the individual group estimates, that is, the standard errors of $B_{0j,\mathrm{OLS}}$ and $B_{1j,\mathrm{OLS}}$. The more precise the estimates $B_{0j,\mathrm{OLS}}$ and $B_{1j,\mathrm{OLS}}$, the more we are willing to rely on the coefficients derived for the individual group. Yet, at the same time, it is advantageous to capitalize on the highly precise estimates based on the whole data set, $B_{0j,\mathrm{POP}}$ and $B_{1j,\mathrm{POP}}$. Because the empirical Bayes (EB) estimators use the information taken from the full sample to estimate coefficients for each individual group, the EB estimates for individual groups are said to "borrow strength" from the estimates based on the whole data set. As indicated, the EB estimators are also called *shrinkage estimators*, because the estimates for individual group coefficients are drawn to (shrink toward) the overall population estimates. The term *shrinkage* as used here is completely unrelated to the use of the term *shrinkage* in the context of unbiased estimates of the squared multiple correlation, described in Section 3.5.3.

The relative weighting of the two sets of estimates in deriving the compromise EB estimate depends on the precision of the estimates from the individual group j. The following is an expression for the EB estimator $\underline{B_{0j,\text{EB}}}$ of the level 1 regression intercept in group j:

$$\underline{B_{0j,\text{EB}}} = \lambda_{0j}\,\underline{B_{0j,\text{OLS}}} + (1 - \lambda_{0j})\,\underline{B_{0j,\text{POP}}}. \tag{14.10.3}$$

The weight λ_{0j}, which is the measure of stability of the OLS coefficient $\underline{B_{0j}}$ for group j, ranges from 0 to 1 and varies inversely as the size of the squared standard error $SE^2_{0j,\text{OLS}}$ of the OLS coefficient $\underline{B_{0j,\text{OLS}}}$ from the sample:

$$\lambda_{0j} = \frac{\tau_{00}}{\tau_{00} + SE^2_{B_{0j,\text{OLS}}}}. \tag{14.10.4}$$

As can be seen from Eq. (14.10.3), the lower the precision of the estimate of the measure on the group (i.e., the smaller λ_{0j}), the more the EB estimator $\underline{B_{0j,\text{EB}}}$ of the random intercept for that group is drawn to the estimator based on the population fixed effects $\underline{B_{0j,\text{POP}}}$, from Eq. (14.10.1). An analogous shrinkage estimator of the level 1 regression slope in group j is given as follows:

$$\underline{B_{1j,\text{EB}}} = \lambda_{1j}\,\underline{B_{1j,\text{OLS}}} + (1 - \lambda_{1j})\,\underline{B_{1j,\text{POP}}}. \tag{14.10.5}$$

Numerical Example: Empirical Bayes Estimation

A numerical example, based on the analysis in Table 14.9.1A, illustrates the EB estimates. Here the variable W in Eq. (14.10.1) and (14.10.2) is TREATC, the treatment condition. We consider a group assigned to the control condition. In the control condition, the value of the level 2 TREATC predictor is $-.596$. Using Eq. (14.10.1) and (14.10.2), we obtain the estimates from the full data set:

$$\underline{B_{0j,\text{POP}}} = \hat{\gamma}_{00} + \hat{\gamma}_{01}W_j = 15.166 + 1.528(-.596) = 14.245;$$

$$\underline{B_{1j,\text{POP}}} = \hat{\gamma}_{10} + \hat{\gamma}_{11}W_j = 3.130 + 1.245(-.596) = 3.130 - .742 = 2.388.$$

In an OLS regression analysis of a single group in the control condition with $n = 9$ cases, we obtain the OLS estimates $\underline{B_{0j,\text{OLS}}} = 15.626$ and $\underline{B_{1j,\text{OLS}}} = 4.750$. There is quite a discrepancy between the estimate of the relationship between motivation and weight loss in the single sample versus the overall data set. In the overall data set, it is estimated that there is a loss of 2.388 pounds for each 1-unit increase in motivation; in the single sample the estimate of pounds lost as a function of motivation is twice as high at 4.750 pounds. However, the standard error of $\underline{B_{1j,\text{OLS}}}$ is quite large, $SE^2_{1j,\text{OLS}} = 1.475$. The weight for $\underline{B_{1j,\text{OLS}}}$ in the EB estimator is

$$\lambda_{1j} = \frac{\tau_{11}}{\tau_{11} + SE^2_{B_{1j,\text{OLS}}}} = \frac{.556}{.556 + 1.475} = .274. \tag{14.10.6}$$

Then the EB estimate of the slope in the group is

$$\underline{B_{1j,\text{EB}}} = \lambda_{1j}\,\underline{B_{1j,\text{OLS}}} + (1 - \lambda_{1j})\,\underline{B_{1j,\text{POP}}} = .274(4.750) + (1 - .274)\,(2.388) = 2.765.$$

This EB estimate for the slope in the single group has moved much closer to (has "shrunken" to) the overall estimate of the slope from the full data set.

For the intercept, the squared standard error $SE^2_{0j,\text{OLS}} = 1.096$. The value of $\lambda_{0j} = 1.967/(1.967 + 1.096) = .642$. Then the EB estimate of the intercept for the group is

$$\underline{B_{0j,\text{EB}}} = \lambda_{0j}\, \underline{B_{0j,\text{OLS}}} + (1 - \lambda_{0j})\, \underline{B_{0j,\text{POP}}} = .642\,(15.626) + (1 - .642)(14.245) = 15.131.$$

Finally, the EB estimate of the regression equation for group j is

$$\underline{\hat{Y}_{ij,\text{EB}}} = 2.765 \text{ MOTIVATC} + 15.131.$$

This equation is a compromise between the OLS regression equation for the sample, based on only 9 cases and with a large standard error for the sample slope $\underline{B_{1j,\text{OLS}}}$, and the overall equation from the full data set, $\underline{\hat{Y}_{ij,\text{POP}}} = 2.388$ MOTIVATC + 14.245.

14.11 STATISTICAL TESTS IN MULTILEVEL MODELS

We have examined tests of significance of both fixed and random effects in the numerical example in Tables 14.5.1 and 14.5.2. Here we provide more information about the tests themselves, following expositions by Raudenbush and Bryk (2002) and Singer (1998).

14.11.1 Fixed Effects

Tests of the fixed effects are made against the standard error of the fixed effect, resulting in a z test. Alternatively, a t test is computed, as is given in both SAS PROC MIXED and the specialized multilevel software package HLM (Raudenbush, Bryk, Cheong, & Congdon, 2001). Degrees of freedom for the test depend on whether the predictor is a level 2 predictor or a level 1 predictor. For level 1 predictors, the df depend on the numbers of individual cases, groups, and level 1 predictors. For level 2 predictors, the df depend on number of groups and number of level 2 predictors, and are specifically $(g - S_q - 1)$ df, where g is the number of contexts (groups) and S_q is the number of level 2 predictors.

14.11.2 Variance Components

Each variance component may be tested for significance of difference from zero in one of several ways. First is a chi square test, based on OLS estimates of within group coefficients, which contrasts within group estimates with the fixed population estimate. Use of this test requires that most or all contexts be of sufficient size to yield OLS estimates. The result is distributed approximately as χ^2, with $(g - S_q - 1)$ df, where g is the number of contexts (groups) from which OLS estimates can be obtained and S_q is the number of level 2 predictors (this test is reported in HLM output). Second is a z test based on large sample theory, reported in SAS PROC MIXED. Both Raudenbush and Bryk (2002) and Singer (1998) express caution concerning this latter test because of the skew of the sampling distribution of the variance components and because of the dependence on large sample theory (asymptotic normality is assumed but not achieved). There is a third approach to examining variance components, a *model comparison approach*, based on likelihood ratio tests of nested models. This is the same form of test as in the testing of nested models in logistic regression explained in Section 13.2.14. In the RC context we specify a model that allows a particular variance component to be nonzero, for example, the variance of the slopes. We then specify a second, more restrictive, model that forces this variance component to zero. A likelihood ratio χ^2 test is used to test whether the model fit is significantly worse when the variance component is forced to zero. If so, we conclude that the variance component is nonzero.

14.12 SOME MODEL SPECIFICATION ISSUES

There are many issues in the specification and execution of multilevel models. We address two related issues here: (1) the issue of whether there are instances in which the same variable can serve as a level 1 and level 2 predictor in a single equation, and (2) issue of centering variables in multilevel models.

14.12.1 The Same Variable at Two Levels

There is a very interesting but theoretically complex possibility that the same variable can exist at more than one level in a single data set. A classic example from education is the impact on academic achievement of the socioeconomic status (SES) of the child versus the average SES of the children in the child's school (Burstein, 1980; see also Raudenbush & Bryk, 2002). In a group therapy context we might measure the depression level of an individual client at the outset of therapy versus the average depression level among all members in the group; we might hypothesize that being in a group of very depressed individuals would impede the progress of an individual client in overcoming his or her own depression. The conceptual issue in both these examples is whether we can provide a distinct theoretical role for the same variable at the group level and the individual level. For the therapy setting we might argue that the average level of depression in the group reflects the depressed cognitions expressed by group members in response to statements by the individual client in question, whereas the client's own level of depression at therapy outset would reflect the level of individual disturbance. These two aspects of depression might have separate influences or even interact in predicting therapy outcome.

14.12.2 Centering in Multilevel Models

Centering in multilevel models is useful for decreasing multicollinearity among predictors and between random intercepts and slopes, thereby stabilizing the analysis. Centering in multilevel models is more complex than in OLS regression. Kreft and de Leeuw (1998) provide a straightforward exposition; a more technical exposition is given in Kreft, de Leeuw, and Aiken (1995). Singer (1998) provides advice on centering in contextual models in the multilevel framework.

For level 1 variables there are two common options for centering: (1) centering each score around the grand mean (CGM) of all the cases in the sample, ignoring group membership, and (2) centering each score around the group mean in which the case occurs, referred to as centering within context (CWC) (Kreft, de Leeuw, & Aiken, 1995). We contrast these approaches with retaining data in their raw score (RS) form.

The CGM approach simply involves subtracting a single constant from each score in the whole distribution regardless of group. In a multilevel model containing a cross-level interaction, the resulting fixed effect parameters change from raw score (RS) to CGM scaling, as does the variance component of the intercept, but there is a straightforward algebraic relationship between the CGM and RS results. It is parallel to the algebraic relationship shown in Section 7.2.5 for centering OLS in equations including interactions. Moreover, measures of fit, predicted scores, and residual scores remain the same across RS and CGM.

CWC is a very different matter. CWC eliminates all the information on mean level differences between the groups because the group means are subtracted out. The mean of each group becomes zero. In order to not lose valuable information (in our weight example, information on the mean pounds lost per group), one must build back mean weight loss per group as a level 2 variable. We refer to the use of CWC without building back group means as CWC_1 and CWC with the group means entered at the second level as CWC_2 in Kreft, de Leeuw, and Aiken

(1995). With CWC_1 there is no way to recover the between class information available in the RS analysis. Only if it can be powerfully argued that differences between the group means on a predictor bear no relationship to the outcome would one consider using CWC_1. CWC_2 reinstates the eliminated between class mean differences at level 2. Even if the focus is on the impact of within group variation on a predictor, we recommend CWC_2; then the presence versus absence of effect of a predictor at the group level becomes a matter for exploration, rather than an untested assumption (see Singer, 1998, for an example).

The choice of centering approach depends on the research question. CGM was used in all the multilevel numerical examples reported here. The rationale for this choice was that the level 1 motivation variable was a person variable to be controlled when the impact of the intervention was assessed; there was no theoretical rationale provided for a special role of the mean level of motivation in the group versus the motivation of the individual within the group to which she belonged on dieting outcomes. Had there been such an interest, we would have used the CWC_2 approach for characterizing motivation and then added the treatment variable as a level 2 predictor. We might have hypothesized a level 2 interaction between treatment and mean level of motivation of the group as well as, or instead of, the cross-level interaction between treatment and individual level of motivation.

Level 2 variables may be centered or not. We also centered the level 2 treatment predictor of treatment (experimental, control) around the grand mean, using weighted effect codes (Chapter 8) at the individual level. We did so to avoid multicollinearity and estimation difficulties, and to facilitate interpretation. (See footnote 5.)

Users of specialized software for multilevel modeling (mentioned in Section 14.15) should take caution to understand how data are being centered by the software. A safe approach is to first center (or not) the level 1 and level 2 data in the form assumed appropriate for the particular problem at hand and then enter the data into the software. This assures that one will obtain the centering desired.

14.13 STATISTICAL POWER OF MULTILEVEL MODELS

OLS ignoring group membership has a substantially inflated Type I error rate when group sizes are large and the intraclass correlation increases. Kreft and de Leeuw (1998) discuss the complex issues involved in statistical power for multilevel models and provide a summary of simulation studies on power. Statistical power must be addressed separately for level 1 and level 2 effects. Power for level 2 effects is dependent on number of groups, power for level 1, on number of cases. Simulation studies suggest that large samples are needed for adequate power in multilevel models and that the number of groups is more important than the number of cases per group. Kreft and de Leeuw suggest that at least 20 groups are needed to detect cross-level interactions when group sizes are not too small. The whole issue of statistical power is complicated, because the power differs for fixed effects versus random effects as a function of effect size and intraclass correlation, and both the number of groups and number of cases per group.

14.14 CHOOSING BETWEEN THE FIXED EFFECTS MODEL AND THE RANDOM COEFFICIENT MODEL

We began the discussion of the handling of clustered data with an exposition of OLS approaches, among which was a *fixed effects model* in which the dependent variable Y was predicted from level 1 variables plus a set of $(g - 1)$ dummy codes to account for differences in means of the g groups or clusters in the data. The alternative presented was the *random coefficient (RC) regression model* in Eqs. (14.4.3) through Eq. (14.6.6). Snijders and Bosker (1999) provide a clear exposition of the issues in choosing and provide recommendations, several of which are mentioned here. The choice depends on the number of groups, the sample sizes per group, the distribution of the level 1 and level 2 residuals, assumptions about how groups were sampled, the resulting generalizations one wishes to make, and the focus of the analysis. Constraints on the available data may dictate choice. With a small number of groups (fewer than 10 according to Snijders and Bosker, 1999), the fixed effects approach is recommended. Further, if the individual groups have special meaning (e.g., various ethnicities) and one wishes to speak to the model within each of the special groups, then the fixed effects approach is more appropriate. On the other hand, if the groups are merely a random sample from a larger population of groups and one wishes to generalize to the population of groups (e.g., from a sample of families to the population of families), then the RC approach is appropriate. Small numbers of cases per group lead one to the RC model, since shrinkage estimators can borrow strength from the full data set and the fixed effects approach may be very unstable (with large standard errors). On the other hand the RC model as it is typically implemented makes the assumption that the level 2 residuals are normally distributed.

14.15 SOURCES ON MULTILEVEL MODELING

We have barely scratched the surface of the large and complex area of multilevel modeling. As previously pointed out, Kreft and de Leeuw (1998) is an excellent starting point for further studying of multilevel models, followed by Snijders and Bosker (1999), and then by Raudenbush and Bryk (2002) and finally by Goldstein (1995). Singer (1998) provides a highly accessible introduction to the use of SAS PROC MIXED for random coefficient regression. Littell, Milliken, Stroup, and Wolfinger (1996) provide extensive documentation of SAS PROC MIXED. In addition to SAS PROC MIXED, there are several specialized software packages for multilevel modeling, of which the HLM software (Raudenbush, Bryk, Cheong, & Congdon, 2001) is perhaps the easiest to use. MlwiN (Goldstein et al., 1998) is another popular package. Chapter 15 of Snijders and Bosker (1999) is devoted to a review of software that can accomplish multilevel models.

14.16 MULTILEVEL MODELS APPLIED TO REPEATED MEASURES DATA

This section has addressed only clustering of individuals within groups. However, RC regression and multilevel models can also be applied to clustering (or serial dependency) that results from having repeated measurements on individuals. In repeated measures applications, the individual becomes the level 2 unit of aggregation. The repeated observations on each individual become the level 1 units. The interest in such repeated measures applications is on modeling an overall trajectory of how individuals on average change over time (the population fixed effects), of developing a trajectory for each individual, and in modeling the individual differences in the trajectory from both level 1 and level 2 predictors. Repeated measures multilevel analysis is presented in Section 15.4. In addition, Snijders and Bosker (1999) provide an accessible introduction.

14.17 SUMMARY

OLS regression and regression approaches subsumed under the generalized linear model (Chapter 13) all assume that observations are independent. When data are clustered such that individuals within clusters (e.g., children within the same family) are more like one another than are randomly selected individuals, bias is introduced into inference in OLS regression. Standard errors of OLS regression coefficients are negatively biased (too small), leading to alpha inflation (Section 14.1). Degree of clustering is measured with the intraclass correlation (ICC, Section 14.1). Clustering can be handled within the OLS regression framework by adding code predictors that identify the clusters and account for cluster differences (Section 14.2). A newer alternative regression model, the random coefficient (RC) regression analysis, handles clustering in a different way from OLS. RC regression permits the appropriate modeling of the impact of individual level predictors on a DV when data are clustered, yielding proper estimates of standard errors (Section 14.3). The RC regression model is presented, and its components explained: level 1 and level 2 equations, the mixed model equation, the variance components (Section 14.4). Concepts of fixed versus random parts of the model are explained. A numerical example is provided (Section 14.5). Clustering in data can be a meaningful aspect of the data (Section 14.6). Random coefficient regression is used to model multilevel data (Section 14.6), that is, data that contain predictors measured at different levels of aggregation, for example, on individual children within families (level 1) and on the families themselves (level 2). The multilevel RC regression model is developed (Section 14.7) with predictors at two levels of aggregation. The use of the multilevel model for the analysis of experiments in which individual differences may interact with treatment effects is developed (Section 14.8) and illustrated with a numerical example (Section 14.9). Estimation of fixed effects, variance components and empirical Bayes estimators are explained (Section 14.10). Statistical tests of the fixed and random components of RC regression are described (Section 14.11). Issues in model specification, including centering of predictors are addressed (Section 14.12). The choice between OLS regression based approaches and the random coefficient approach to handling clustered data is discussed (Section 14.14).