In Chapter 8 we learned about the concept of hierarchical modeling, a data analysis approach that is appropriate when we have multiple measurements within each of several groups. In that chapter, variation in the data was represented with a between-group sampling model for group-specific means, in addition to a within-group sampling model to represent heterogeneity of observations within a group. In this chapter we extend the hierarchical model to describe how relationships between variables may differ between groups. This can be done with a regression model to describe within-group variation, and a multivariate normal model to describe heterogeneity among regression coefficients across the groups. We also cover estimation for hierarchical generalized linear models, which are hierarchical models that have a generalized linear regression model representing within-group heterogeneity.

11.1 A hierarchical regression model

Let’s return to the math score data described in Section 8.4, which included math scores of 10th grade children from 100 different large urban public high schools. In Chapter 8 we estimated school-specific expected math scores, as well as how these expected values varied from school to school. Now suppose we are interested in examining the relationship between math score and another variable, socioeconomic status (SES), which was calculated from parental income and education levels for each student in the dataset.

In Chapter 8 we quantified the between-school heterogeneity in expected math score with a hierarchical model. Given the amount of variation we observed it seems possible that the relationship between math score and SES might vary from school to school as well. A quick and easy way to assess this possibility is to fit a linear regression model of math score as a function of SES for each of the 100 schools in the dataset. To make the parameters more interpretable we will center the SES scores within each school separately, so that the sample average SES score within each school is zero. As a result, the
intercept of the regression line can be interpreted as the school-level average math score.

![Graph showing least squares regression lines for the math score data, and plots of estimates versus group sample size.](image)

**Fig. 11.1.** Least squares regression lines for the math score data, and plots of estimates versus group sample size.

The first panel of Figure 11.1 plots least squares estimates of the regression lines for the 100 schools, along with an average of these lines in black. A large majority show an increase in expected math score with increasing SES, although a few show a negative relationship. The second and third panels of the figure relate the least squares estimates to sample size. Notice that schools with the highest sample sizes have regression coefficients that are generally close to the average, whereas schools with extreme coefficients are generally those with low sample sizes. This phenomenon is reminiscent of what we discussed in Section 8.4: The smaller the sample size for the group, the more probable that unrepresentative data are sampled and an extreme least squares estimate is produced. As in Chapter 8, our remedy to this problem will be to stabilize the estimates for small sample size schools by sharing information across groups, using a hierarchical model.

The hierarchical model in the linear regression setting is a conceptually straightforward generalization of the normal hierarchical model from Chapter 8. We use an ordinary regression model to describe within-group heterogeneity of observations, then describe between-group heterogeneity using a sampling model for the group-specific regression parameters. Expressed symbolically, our within-group sampling model is

$$Y_{i,j} = \beta_j^T x_{i,j} + \epsilon_{i,j}, \quad \{\epsilon_{i,j}\} \sim \text{i.i.d. normal}(0, \sigma^2),$$  \hspace{1cm} (11.1)

where $x_{i,j}$ is a $p \times 1$ vector of regressors for observation $i$ in group $j$. Expressing $Y_{1,j}, \ldots, Y_{n_j,j}$ as a vector $Y_j$ and combining $x_{1,j}, \ldots, x_{n_j,j}$ into an $n_j \times p$ matrix $X_j$, the within-group sampling model can be expressed equivalently as $Y_j \sim \text{multivariate normal}(X_j \beta_j, \sigma^2 I)$, with the group-specific data vectors $Y_1, \ldots, Y_m$ being conditionally independent given $\beta_1, \ldots, \beta_m$ and $\sigma^2$. 