

```

R code  a.hat.M1 <- coef(M1)$county[,1]    # 1st column is the intercept
        b.hat.M1 <- coef(M1)$county[,2]    # 2nd element is the slope
        x.jitter <- x + runif(n,-.05,.05)  # jittered data for plotting
        par (mfrow=c(2,4))                 # make a 2x4 grid of plots
        for (j in display8){
          plot (x.jitter[county==j], y[county==j], xlim=c(-.05,1.05),
                ylim=y.range, xlab="floor", ylab="log radon level", main=uniq.name[j])
        ## [uniq.name is a vector of county names that was created earlier]
          curve (coef(lm.pooled)[1] + coef(lm.pooled)[2]*x, lty=2, col="gray10",
                add=TRUE)
          curve (coef(lm.unpooled)[j+1] + coef(lm.unpooled)[1]*x, col="gray10",
                add=TRUE)
          curve (a.hat.M1[j] + b.hat.M1[j]*x, lwd=1, col="black", add=TRUE)
        }

```

Here, `lm.pooled` and `lm.unpooled` are the classical regressions that we have already fit.

More complicated models

The `lmer()` function can also handle many of the multilevel regressions discussed in this part of the book, including group-level predictors, varying intercepts and slopes, nested and non-nested structures, and multilevel generalized linear models. Approximate routines such as `lmer()` tend to work well when the sample size and number of groups is moderate to large, as in the radon models. When the number of groups is small, or the model becomes more complicated, it can be useful to switch to Bayesian inference, using the Bugs program, to better account for uncertainty in model fitting. We return to this point in Section 16.1.

12.5 Five ways to write the same model

We begin our treatment of multilevel models with the simplest structures—*nested* models, in which we have observations $i = 1, \dots, n$ clustered in groups $j = 1, \dots, J$, and we wish to model variation among groups. Often, predictors are available at the individual and group levels. We shall use as a running example the home radon analysis described above, using as predictors the house-level x_i and a measure of the logarithm of soil uranium as a county-level predictor, u_j . For some versions of the model, we include these both as individual-level predictors and label them as X_{i1} and X_{i2} .

There are several different ways of writing a multilevel model. Rather than introducing a restrictive uniform notation, we describe these different formulations and explain how they are connected. It is useful to be able to express a model in different ways, partly so that we can recognize the similarities between models that only appear to be different, and partly for computational reasons.

Allowing regression coefficients to vary across groups

Perhaps the simplest way to express a multilevel model generally is by starting with the classical regression model fit to all the data, $y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \epsilon_i$, and then generalizing to allow the coefficients β to vary across groups; thus,

$$y_i = \beta_{0\ j[i]} + \beta_{1\ j[i]} X_{i1} + \beta_{2\ j[i]} X_{i2} + \dots + \epsilon_i.$$

The “multilevel” part of the model involves assigning a multivariate distribution to the vector of β ’s within each group, as we discuss in Section 13.1.

For now we will focus on *varying-intercept models*, in which the only coefficient that varies across groups is the constant term β_0 (which, to minimize subscripting, we label α). For the radon data that include the floor and a county-level uranium predictor, the model then becomes

$$y_i = \alpha_{j[i]} + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i$$

where X_{i1} is the i^{th} element of the vector $X_{(1)}$ representing the first-floor indicators and X_{i2} is the i^{th} element of the vector $X_{(2)}$ representing the uranium measurement in the county containing house i . We can also write this in matrix notation as

$$y_i = \alpha_{j[i]} + X_i \beta + \epsilon_i$$

with the understanding that X includes the first-floor indicator and the county uranium measurement but not the constant term. This is the way that models are built using `lmer()`, including all predictors at the individual level, as we discuss in Section 12.6.

The second level of the model is simply

$$\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2). \quad (12.5)$$

Group-level errors. The model (12.5) can also be written as

$$\alpha_j = \mu_\alpha + \eta_j, \text{ with } \eta_j \sim N(0, \sigma_\alpha^2). \quad (12.6)$$

The group-level errors η_j can be helpful in understanding the model; however, we often use the more compact notation (12.5) to reduce the profusion of notation. (We have also toyed with notation such as $\alpha_j = \mu^\alpha + \epsilon_j^\alpha$ in which ϵ is consistently used for regression errors—but the superscripts seem too confusing. As illustrated in Part 2B of this book, we sometimes use such notation when programming models in Bugs.)

Combining separate local regressions

An alternative way to write the multilevel model is as a linking of local regressions in each group. Within each group j , a regression is performed on the local predictors (in this case, simply the first-floor indicator, x_i), with a constant term α that is indexed by group:

$$\text{within county } j: y_i \sim N(\alpha_j + \beta x_i, \sigma_y^2), \text{ for } i = 1, \dots, n_j. \quad (12.7)$$

The county uranium measurement has not yet entered the model since we are imagining separate regressions fit to each county—there would be no way to estimate the coefficient for a county-level predictor from any of these within-county regressions.

Instead, the county-level uranium level, u_j , is included as a predictor in the second level of the model:

$$\alpha_j \sim N(\gamma_0 + \gamma_1 u_j, \sigma_\alpha^2). \quad (12.8)$$

We can also write the distribution in (12.8) as $N(U_j \gamma, \sigma_\alpha^2)$, where U has two columns: a constant term, $U_{(0)}$, and the county-level uranium measurement, $U_{(1)}$. The errors in this model (with mean 0 and standard deviation σ_α) represent variation *among counties* that is not explained by the local and county-level predictors.

The multilevel model combines the J local regression models (12.7) in two ways: first, the local regression coefficients β are the same in all J models (an assumption we will relax in Section 13.1). Second, the different intercepts α_j are connected through the group-level model (12.8), with consequences to the coefficient estimates that we discuss in Section 12.6.

Group-level errors. We can write (12.8) as

$$\alpha_j = \gamma_0 + \gamma_1 u_j + \eta_j, \quad \text{with } \eta_j \sim N(0, \sigma_\alpha^2), \quad (12.9)$$

explicitly showing the errors in the county-level regression.

Modeling the coefficients of a large regression model

The identical model can be written as a single regression, in which the local and group-level predictors are combined into a single matrix X :

$$y_i \sim N(X_i \beta, \sigma_y^2), \quad (12.10)$$

where, for our example, X includes vectors corresponding to:

- A constant term, $X_{(0)}$;
- The floor where the measurement was taken, $X_{(1)}$;
- The county-level uranium measure, $X_{(2)}$;
- J (not $J-1$) county indicators, $X_{(3)}, \dots, X_{(J+2)}$.

At the upper level of the model, the J county indicators (which in this case are $\beta_3, \dots, \beta_{J+2}$) follow a normal distribution:

$$\beta_j \sim N(0, \sigma_\alpha^2), \quad \text{for } j = 3, \dots, J+2. \quad (12.11)$$

In this case, we have centered the β_j distribution at 0 rather than at an estimated μ_β because any such μ_β would be statistically indistinguishable from the constant term in the regression. We return to this point shortly.

The parameters in the model (12.10)–(12.11) can be identified exactly with those in the separate local regressions above:

- The local predictor x in model (12.7) is the same as $X_{(1)}$ (the floor) here.
- The local errors ϵ_i are the same in the two models.
- The matrix of group-level predictors U in (12.8) is just $X_{(0)}$ here (the constant term) joined with $X_{(2)}$ (the uranium measure).
- The group-level errors $\eta_1, \dots, \eta_J$ in (12.9) are identical to $\beta_3, \dots, \beta_{J+2}$ here.
- The standard-deviation parameters σ_y and σ_α keep the same meanings in the two models.

Moving the constant term around. The multilevel model can be written in yet another equivalent way by moving the constant term:

$$\begin{aligned} y_i &= N(X_i \beta, \sigma_y^2), \quad \text{for } i = 1, \dots, n \\ \beta_j &\sim N(\mu_\alpha, \sigma_\alpha^2), \quad \text{for } j = 3, \dots, J+2. \end{aligned} \quad (12.12)$$

In this version, we have removed the constant term from X (so that it now has only $J+2$ columns) and replaced it by the equivalent term μ_α in the group-level model. The coefficients $\beta_3, \dots, \beta_{J+2}$ for the group indicators are now centered around μ_α rather than 0, and are equivalent to $\alpha_1, \dots, \alpha_J$ as defined earlier, for example, in model (12.9).

Regression with multiple error terms

Another option is to re-express model (12.10), treating the group-indicator coefficients as error terms rather than regression coefficients, in what is often called a

“mixed effects” model popular in the social sciences:

$$\begin{aligned} y_i &\sim N(X_i\beta + \eta_{j[i]}, \sigma_y^2), \text{ for } i = 1, \dots, n \\ \eta_j &\sim N(0, \sigma_\alpha^2), \end{aligned} \quad (12.13)$$

where $j[i]$ represents the county that contains house i , and X now contains only three columns:

- A constant term, $X_{(0)}$;
- The floor, $X_{(1)}$;
- The county-level uranium measure, $X_{(2)}$.

This is the same as model (12.10)–(12.11), simply renaming some of the β_j ’s as η_j ’s. All our tools for multilevel modeling will automatically work for models with multiple error terms.

Large regression with correlated errors

Finally, we can express a multilevel model as a classical regression with correlated errors:

$$y_i = X_i\beta + \epsilon_i^{\text{all}}, \quad \epsilon^{\text{all}} \sim N(0, \Sigma), \quad (12.14)$$

where X is now the matrix with three predictors (the constant term, first-floor indicator, and county-level uranium measure) as in (12.13), but now the errors ϵ_i^{all} have an $n \times n$ covariance matrix Σ . The error ϵ_i^{all} in (12.14) is equivalent to the sum of the two errors, $\eta_{j[i]} + \epsilon_i$, in (12.13). The term $\eta_{j[i]}$, which is the same for all units i in group j , induces correlation in ϵ^{all} .

In multilevel models, Σ is parameterized in some way, and these parameters are estimated from the data. For the nested multilevel model we have been considering here, the variances and covariances of the n elements of ϵ^{all} can be derived in terms of the parameters σ_y and σ_α :

$$\begin{aligned} \text{For any unit } i: \quad \Sigma_{ii} &= \text{var}(\epsilon_i^{\text{all}}) = \sigma_y^2 + \sigma_\alpha^2 \\ \text{For any units } i, k \text{ within the same group } j: \quad \Sigma_{ik} &= \text{cov}(\epsilon_i^{\text{all}}, \epsilon_k^{\text{all}}) = \sigma_\alpha^2 \\ \text{For any units } i, k \text{ in different groups:} \quad \Sigma_{ik} &= \text{cov}(\epsilon_i^{\text{all}}, \epsilon_k^{\text{all}}) = 0. \end{aligned}$$

It can also be helpful to express Σ in terms of standard errors and correlations:

$$\begin{aligned} \text{sd}(\epsilon_i) &= \sqrt{\Sigma_{ii}} = \sqrt{\sigma_y^2 + \sigma_\alpha^2} \\ \text{corr}(\epsilon_i, \epsilon_k) &= \frac{\Sigma_{ik}}{\sqrt{\Sigma_{ii}\Sigma_{kk}}} = \begin{cases} \frac{\sigma_\alpha^2}{\sigma_y^2 + \sigma_\alpha^2} & \text{if } j[i] = j[k] \\ 0 & \text{if } j[i] \neq j[k]. \end{cases} \end{aligned}$$

We generally prefer modeling the multilevel effects explicitly rather than burying them as correlations, but once again it is useful to see how the same model can be written in different ways.

12.6 Group-level predictors

Adding a group-level predictor to improve inference for group coefficients α_j

We continue with the radon example from Sections 12.2–12.3 to illustrate how a multilevel model handles predictors at the group as well as the individual levels.