Chapter 7

Estimation of Mixed Models

This chapter introduces an approach to repeated measures/longitudinal data that is based on a philosophy of estimation that is fundamentally different than the estimating equation approach discussed in the previous chapter. Without going into the gory technical details, the estimating equation approach: 1) is based on specifying a particular parameter of interest (not the entire distribution of the data), 2) deriving general estimating equations that return consistent estimates of these parameters under certain assumptions and 3) return robust inference for these parameter estimates that does not rely on correct specification of other aspects of the distribution of the data (i.e., the precise correlation structure among measurements on the same individual). In contrast, the mixed model is a maximum likelihood procedure that is based on correct specification of both the mean model (the regression) and the "error" distribution - the correlation structure of the residual errors. By mathematical luck (e.g., in case of linear mixed models), sometimes the estimation of the coefficients is not biased even if this correlation structure is mispecified. However, the

inference on the coefficients does rely on getting the correlation structure correct. And, in the logistic case, the mixed model even relies on the precise distributional assumptions regarding the unmeasured random effect (usually, that it is normally distributed). Besides these downsides, the ability to specify a rich variety of variance-covariance models using random effects, random coefficients and directly specifying several covariance structures on the residual error makes the mixed model approach particularly attractive when the goals are twofold: 1) getting estimates of the regression coefficients (the singular goal of the estimating equation approach) and 2) estimating the covariance structure of the data. In addition, one can even derive estimates (technically, predictions) of the random effects for individuals in the data set. In this chapter, we start by discussing the general framework for mixed models. Subsequently we focus on two of the most commonly used of the generalized linear models with longitudinal data: linear and logistic. Though we do not discuss specific examples of log-linear models, such as Poisson, the extrapolation of the use of log-linear models from the information provided in the linear and logistic sections is very straightforward.

Mixed models are a very popular analysis tool, not just for longitudinal studies, but for studies of other correlated observations, such as studies of people in neighborhoods and studies of students in classrooms with many other examples as well. They are part of a larger class of latent variable models that propose the observed set of outcomes for a unit i ($\mathbf{Y}_i$) is a product of both observed covariates, $\mathbf{X}_i$ and also unobserved (latent) variables related to these units (the α_i). Not only this, but that one can specify the form and distribution of how these latent variables are related to $\mathbf{Y}_i$. Other forms of latent variable

models include structural equation models and the subset of such models called confirmatory factor analysis. The strength of these approaches is that they allow a almost infinite number of choices for how to invoke the involvement and specific statistical relationships of unmeasured variables to outcomes of interest (and even unmeasured outcomes to measured proxies of such outcomes). Thus, the variety of how these models are constructed is really only limited by the imagination of the researcher (and the ability of the software to find a solution given the data at hand). That is also their weakness - interpreting the results of these models relies on trusting the often untestable assumptions of the latent variable distributions. If one is ambitious about proposing elaborate models based on untestable assumptions, the gains are potentially great if one is right or perhaps not too wrong. But, then one could be wrong enough so that the results obtained misleading, and the data can not be relied upon to help the researcher know the difference. This is a general issue with latent variable models we discuss at the end of this chapter (section 7.4).

Before diving into the gory details of various kinds of mixed models, a first obvious question is what is being mixed? The deceptively simple answer is two kinds of effects: fixed and random and they refer to two types of coefficients in the models (in our case, linear, log-linear and logistic). Fixed effect coefficients are parameters of the data-generating distribution (e.g., the average association of log(viral load) and CD4 count among HIV+ subjects post-HAART therapy) and random effects are random variables those that only apply to the r*randomly drawn* entities (the association of log(viral load) and CD4 count only among subject i. This proposition that the observed data is a function of average effects (fixed) and variability caused by differences in these effects among subjects (random) along

with proposing correlation structures on the residual error opens up an almost infinite variety of data-generating distributions.

To find the estimates of interest, one typically must propose specific distributions of these random effects (and most commonly and for this chapter, these are assumed to be normal) and thus one adds another set of parameters not return in your typical non-longitudinal, non-mixed model and that: variance-covariance parameters. For instance, one gets not only the average association of CD4 count and log(viral load) one also gets the estimate of the distribution of the variability of this relationship. That is another strength of this approach is that it can go beyond estimate of average relationships, it can also estimate how these relationships vary in a population and sometimes that's the most interesting parameter (for instance, trying to determine the proportion of subjects that will have a negative impact to a drug).

In the following chapter, we first discuss a general framework for mixed models (7.1), though leave the technical issues to other references. We follow with a detailed discussion of linear mixed effects models (7.2.1), and then other nonlinear mixed effects models, specifically logit and log-linear (**??**. Finally, we have a discussion on a general philosophy of estimation involving latent variable models at the end of this chapter (section 7.4).

7.1 General Framework for Mixed Models

Most generally, mixed models are latent variable models, in that the hypothesize that the variability in the outcome can be explained by a combination (the "mixed" part) of measured covariates, and unmeasured, latent variables attributable to some randomly drawn unit (e.g., individual). We begin this chapter discusses mixed effects models in general (linear and non-linear). We discuss the case where there is a simple hierarchical structure: repeated measures on individual units (e.g., people).

7.1.1 The Likelihood

First, assume there is no covariates, just the outcome, Y_{ij}, $i = 1, \ldots, m$ and $j = 1, \ldots, n_i$. Thus, we have m i.i.d. observations of $\mathbf{Y}_i$, so the likelihood of the data will be the product of the likelihoods of each observations. We will assume the underlying *mixed* model for Y consists of a random effect at the individual level, say α_i and perhaps other *fixed* components, β. We will assume, as is often done, that the observations on an individual are independent *given* the (unobserved) random effects, α_i, or

$$f(\mathbf{Y}_i, \alpha_i) = f(\mathbf{Y}_i \mid \alpha_i) f(\alpha_i),$$

where the conditional independence implies

$$f(\mathbf{Y}_i \mid \alpha_i) = f(Y_{i1} \mid \alpha_i) f(Y_{i2} \mid \alpha_i) \cdots f(Y_{in_i} \mid \alpha_i).$$

Ignoring the mean for now, we will assume the distribution of the data is indexed by a set of variance parameters for the distribution of $\mathbf{Y}_i \mid \alpha_i$, ϕ, and a set of variance parameters

for the model of the distribution of α_i, γ. The likelihood of the observed data then requires us to integrate out the unobserved (latent) random effect variables (intuitively, sense we do not the α_i, to determine the likelihood of a the vector of outcomes for an individual, we sum up the probabilities of the outcome given each possible value of their α_i, where this probabilities are determined by the parameters ϕ, γ):

$$L(\beta, \phi, \gamma) = \prod_{i=1}^{m} \int f_{\phi}(\mathbf{Y}_i \mid \alpha_i) f_{\gamma}(\alpha_i) d\alpha_i.$$

Now, assume we have a model the relates covariates, say $\mathbf{X}_i$ (a matrix of covariate values, where each row j are covariates measured at observation j, to the distribution of the data), via a set of parameters, β (e.g., coefficients). The informative part of the likelihood becomes,

$$L(\beta, \phi, \gamma) = \prod_{i=1}^{m} \int f_{\beta,\phi}(\mathbf{Y}_i \mid \alpha_i, \mathbf{X}_i) f_{\gamma}(\alpha_i \mid \mathbf{X}_i) d\alpha_i. \tag{7.1}$$

In general one estimates the regression coefficients (related to the conditional mean model) and the covariance parameters implied by the random effects model using either maximum likelihood (ML) or so-called restricted maximum likelihood (REML - for details on these approaches see [**?**]). The two techniques will be asymptotically equivalent in their estimates of the variance components, but the difference between ML estimates and REML estimates can important if the sample size is relatively small. Given a correctly specified model, the REML estimates of the variance components are less biased than the corresponding ML estimates, which is illustrated in the simplest example, the sample variance. In a normal model (for say an i.i.d. sample of Y_i, with no repeated measures), the ML estimate of

variance of n i.i.d. observations Y_i is:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

which is biased low. The sample variance, which is a REML estimate of the variance, is:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2,$$

which is unbiased. Also, the asymptotic equivalence is obvious here, because as n gets large, this difference becomes negligible. This will be also true in general of the contrasts of ML and REML covariance parameter estimates in mixed models - as the number of sub-units gets larger (the n_i) the difference will become unimportant.

In general, mixed models, as different from using a GEE approach, are not targeted at a specific part of the distribution (i.e., the mean of the outcome given covariates), but based on defining the likelihood, that is entire distribution of the data. Thus, unlike GEE, they will use ML (or REML) estimation to derive the estimates of the parameters of interest. For linear models, the derivation of the estimates is relatively straightforward. More generally, it can be computationally challenging to find the solutions to these maximum likelihood equations, and so much of the challenge of implementation can be computational. For example, a challenge can be to find more accurate and quicker methods of numerical integration because the integral in equation (7.1) is not typically analytically solvable. In addition, much more could be said technically about how to construct likelihoods based on REML, but we leave that for other sources (ADD REFERENCES). Our simple message is

1) because the implied quantities of interest are the entire distribution, not just the mean, then one has to be more concerned about invoking stronger statistical assumptions when deriving inference, and 2) it's generally better to use REML (though bias is of course only one source of error in a parameter estimate), but if the data is based on a reasonable sample size, the difference will be of no consequence.

Finally, the standard errors (SEs) for the estimated parameters will be, by default, those based on the likelihood, and thus they will typically be sensitive to assumptions of the model (for instance, that the variance model implied by the mixed effects model is correct). However, for some circumstances, one can also derive SEs that are robust to misspecification of the model, much like discussed in the GEE chapter. For instance, if the data-generating distribution implies independent sets of observation by some unit (e.g., individual, school, etc.), then the so-called clustered bootstrap (REF), where the highest level units are randomly sampled with replacement, can be an alternative method for estimating the sampling variability of the parameter estimate. In addition, for the linear model case, there is also the availability of the so-called sandwich estimator (REF), which also is robust to misspecification of the mixed effects model. As we will discuss below, if the model is random effects portion of the model is misspecified, then it is not necessarily straightforward to interpret the resulting parameter estimates. That means that one might get trustworthy inference, but on a parameter that has an unknown interpretation. However, at least having a more nonparametric form of deriving the inference is a good check on the assumptions of the mixed model. If there are big differences in the inferences made using the parametric, likelihood-based (naive) versus nonparametric robust inference, that

should cause some serious concern about the modeling assumptions. In examples below, we report both the naive and robust inference for some of the examples.

7.2 Continuous Response/Linear Mixed Effects Models

As the name implies, a linear mixed effects model is a mixture of parameters/covariates and random effect error terms, or:

$$Y_{ij} = \mathbf{X}_{ij}^T\beta + \mathbf{Z_{ij}}^T\beta_{\mathbf{i}} + e_{ij} \tag{7.2}$$

where Z_ij is some subset of the X_{ij} and β_i is the corresponding set of random effects. To define the relevant part of the likelihood, one thus needs to define the joint distribution of the β_i, and in almost all cases, that is assumed to be multivariate normal with mean 0, or $\beta_i \sim MVN(0, G)$, where G is the user-specified variance covariance matrix. In addition, one also generally assumes that the vector of errors within a unit, i, or e_i is another multivariate normal, or $e_i \sim MVN(0, \Sigma_i)$; for example, one might assume independent, errors drawn from distribution with equal variance, or , $\Sigma_i = \sigma_e^2\mathbf{I}_{n_i x n_i}$. In this case, then the likelihood of an observation on a subject, or $O_i = (\mathbf{Y_i}, X_i)$ is multivariate normal, or $\mathbf{Y_i} \sim MVN(\mathbf{X}_{ij}^T\beta, V_i)$, where the variance-covariance of observations on the same unit, i, is $V_i = \mathbf{Z_{ij}}G\mathbf{Z_{ij}}^T + \Sigma_i$. Thus, the likelihood (conditional distribution of $\mathbf{Y_i} \mid X_i$) in this very particular, but most commonly used linear mixed model is multivariate normal, where the conditional mean is a linear model, and the variance-covariance is a function of the random effects model, and the assumptions on the joint residual error distribution

within unit i. Different random effect models, and different assumptions on the correlation of errors within a unit will result in different overall variance-covariance models for the vector of outcomes on an individual unit. The general formulation is important, but we will now explore specific examples to develop intuition for how different random effect models generate different models for the variance covariance of the outcomes.

7.2.1 Two Levels (observations within subjects), No Covariates

To introduce this section, we will first introduce the simplest of mixed models and show by way of this introduction the many issues that are raised, even by this simple model. Consider the following random intercept model with no explanatory variables:

$$Y_{ij} = \beta_0 + \beta_{0i} + e_{ij}, \tag{7.3}$$

where β_{0i} and e_{ij} are unmeasured random variables. We will make the typical assumption that $E(\beta_{0i}) = E(e_{ij}) = 0$ and these random variables are independent of each other, $\beta_{0i} \perp E(e_{ij})$. In addition, for now we will only describe the distribution of the random variables (apart from their fixed means) by their variances, say $var(\beta_{0i}) = \sigma^2_{\beta_0}$ and $var(e_{ij}) = \sigma^2_e$. Note, that the random variable, β_{0i} only has the index, i, and the e_{ij} has both i and j as indices implying a unique random error makes up part of each of observation on an individual. When the i's are different people, we typically assume that the $\beta_{0i} \perp \beta_{0i'}, i \neq i'$ (these random effects are independent of one another). One can imagine situations where one would wish to relax that assumption (e.g., if people are nested in neighborhoods). In addition, it might make sense to include in the model correlation that diminishes in distance

in time, and we will discuss situations below both where we allow nested units as well as residual correlation in the error terms. For now, assume $e_{ij} \perp e_{ij'}, j \neq j'$.

The first step is to examine whether this model implies correlation of observations on the same individual (the same i). One reason for the popularity of the mixed model approach for repeated measures data is that the answer is yes. One can easily show in this model, $cor(Y_{ij}, Y_{ij'}) = \sigma^2_{\beta_0}/(\sigma^2_{\beta_0} + \sigma^2_e), j \neq j'$, or the correlation is the proportion of the total variance due to the between unit variance (quantity sometimes called the intraclass correlation coefficient). The typical approach to estimation using mixed models is to not start with the desired covariance model, such as the one implied by this simple model, but to invoke the mixed model that is consistent with the analyst's understanding of the sources of variation and co-variation in the data - this will imply a certain variance-covariance model. One can immediately see the contrast with the approach we discussed in Chapter 6 where a variance-covariance model was directly asserted - one can fit a mixed model without ever explicitly defining correlation of observations on the same unit, because that is done implicitly by the chosen model. Further on in this chapter we will present an ad hoc method for choosing among competing mixed models, but for now we will assume that this is already known or assumed.

As an illustration of this method, consider the same CD4 data used in table 6.3. In this case, we fit the model 7.3, which results in table

Table 7.1: Random effects model 7.3 for CD4 of HIV patients under HAART

Parameter	Estimate	SE of Estimate
β_0	315.4	19.9
σ_{β_0}	163.4	16.4
σ_e	125.6	9.5
ρ	0.63	

ADD EXAMPLE AND ALSO CONNECT THE MODEL TO THE GENERAL MODEL ABOVE

7.2.2 Two-Levels with Covariates

Know consider a more complicated version of 7.3 or:

$$Y_{ij} = \beta_0 + \beta_{0i} + \beta_1 X_{ij} + e_{ij}, \tag{7.4}$$

In this case, the same variance and covariance model is assumed as in 7.3 once we have conditioned on X_ij, however, what is the interpretation of β_1 in this model? To find out takes a bit of simple algebra and basically involves comparing two means conditioned on X_ij and keeping the subject (the i) fixed, or

$$E[Y_{ij} \mid X_{ij} = x+1, \beta_{0i}] - E[Y_{ij} \mid X_{ij} = x, \beta_{0i}] = \beta_0 + \beta_{0i} + \beta_1(x+1) - [\beta_0 + \beta_{0i} + \beta_1 x] = \beta_1. \tag{7.5}$$

Thus, it is interpreted as the mean difference within a subject if at two levels of X_{ij} that differ by 1, more vaguely as the subject-specific association of Y_{ij} and X_{ij}. In addition, obviously we also have that if we average over the distribute of β_{0i} in the population, we

Table 7.2: Random effects model 7.4 for CD4 versus binary viral load of HIV patients under HAART

Parameter	Estimate	SE of Estimate
β_0	355.1	21.8
β_1	-79.3	17.1
σ_{β_0}	169.1	16.0
σ_e	113.0	8.6
ρ	0.69	

get the same parameter, or:

$$E_{\beta_{0i}}\{E[Y_{ij} \mid X_{ij} = x + 1] - E[Y_{ij} \mid X_{ij} = x]\} = \beta_1,$$

so that in the case of the linear mixed model, this coefficient has both a "conditional" (on the individual) and "marginal" (averaged over all individual) interpretation.

Based upon this simple example, we examine the HIV CD4 versus viral load data set following a set of analyses done in the previous chapter (see Table **??**).

One can see the results are very similar to the use of exchangeable working correlation matrix with naive inference as shown in table 6.3. This is because both of these have essentially the same estimates, that is from weighted least squares (6.18), where the weight matrix is the inverse of the estimated variance-covariance matrix assumed to be exchangeable. The difference being that this form of the variance-covariance is explicitly defined via the GEE approach (that is, a function of the working correlation and specific statistical model) whereas in the random effects case, it is implicitly defined via the specific

model (7.4). In addition, given the variance estimates of the coefficient estimate in the random effects case also depends on the random effects specification only through the implied variance-covariance model (and the procedure assumes this model is correctly specified) the inference should be nearly the same as the GEE approach with naive inference, and that is verified by comparing the results (the small differences involve small differences in the estimates of the variance parameters).

Now, let's spend some time interpreting these parameters. First of all, β_0 is interpreted as the average intercept among all subjects in the target population, or equivalently, the mean of CD4 count at low viral load. The estimated standard deviation of the intercepts is $\sigma_{\beta_0} = 169.1$. If one takes the normality assumption seriously, one can use this estimate to examine the entire distribution of characteristic CD4 counts for observations with high viral load (or $X_{ij} = 1$). First, to get the mean across observations, we want $\hat{E}[Y_{ij} \mid X_{ij} = 1] = 355.1 - 79.3 = 275.8$. One way to characterize the distribution is the interquartile range, or the estimate 25th and 75th percentile. In this case, the IQR of the distribution $N(\hat{E}[Y_{ij} \mid X_{ij} = 1], \hat{\sigma}_{\beta_0}) = N(275.8, 169.1)$, which is $275.8 \pm 0.67 * 169.1 = (161.7, 389.9)$, where 0.67 is the 75th quantile of the standard normal distribution. Thus, using the estimated model (not just the mean, but entire distribution), the results suggest that around the central 50 % of subjects would, if they had high viral load, have CD4 counts from around 150 to 400. This demonstrates that the amount of apparent information returned by a mixed model is potentially much more than just be a procedure that models the conditional mean, but does not provide estimates of the entire distribution. However, in order to put much faith in distributional estimates, like quantiles in this case, one also

relies strongly on the normality assumption. Given we know that the outcomes are never perfectly normally distributed, it is unclear how seriously one should take such estimates, and thus how useful the additional information provided by mixed models (say relative to GEE) really is. We explore this comparison in more detail at the end of the chapter.

Random Coefficients

Random intercepts are not the only way to add variability due to a statistical unit; the model can also be expanded to include variability across other coefficients across units. For instance, continuing the example started above.

$$Y_{ij} = \beta_0 + \beta_{0i} + (\beta_1 + \beta_{1i})X_{ij} + e_{ij}, \tag{7.6}$$

where we now assume

$beta_{1i} \sim N(0, \sigma_{\beta_1})$. Obviously, we have now introduce the idea that not only the intercept can be random but slopes can be random as well. Also, given there will be no compelling reason to believe the slope intercept are independent, one typically allows these random steps may covary, $\sigma_{\beta_0,\beta_1} \equiv cov(\beta_{0i}, \beta_{1i})$. This results and even more flexibility in defining the variance model for the outcomes conditional on covariates. For instance, note that the model (7.6) implies:

$$var(Y_{ij} \mid X_{ij}) = \sigma_e^2 + \sigma_{\beta_0}^2 + X_{ij}^2\sigma_{\beta_1}^2 + 2X_{ij}\sigma_{\beta_0,\beta_1}.$$

This in turn allows for non-constant variance, as it now depends on X_{ij}.

Table 7.3: Random coefficient model 7.6 for CD4 versus binary viral load of HIV patients under HAART

Parameter	Estimate	SE of Estimate
β_0	355.1	23.5
β_1	-79.3	17.0
σ_{β_0}	207.4	1246.6
σ_{β_1}	123.1	4200.7
$corr(\beta_0, \beta_1)$	-0.58	2.92
σ_e	71.1	3638.5

Table 7.3 shows the estimates from model 7.6, which have a couple of additional variance parameters relative to the results in Table 7.2, specifically an estimate of σ_{β_1} and σ_{β_0,β_1}. One can see no change in the estimates of the coefficients, and very little difference in the standard errors. Though there appears to be substantial difference in the estimates of the variance components, particularly that of σ_{β_0}, this is mainly illusory and the result of the lack of information in the data for distinguishing the contribution of variance due to the random intercepts and slopes. Evidence for the this lack of identifiability comes from the extreme standard errors for the variance components (and though not shown, the strong estimated correlation of the estimates). Thus, though one can try to estimate more complicated mixed models, one can run up against the amount of information available to distinguish amount the sources of variation.

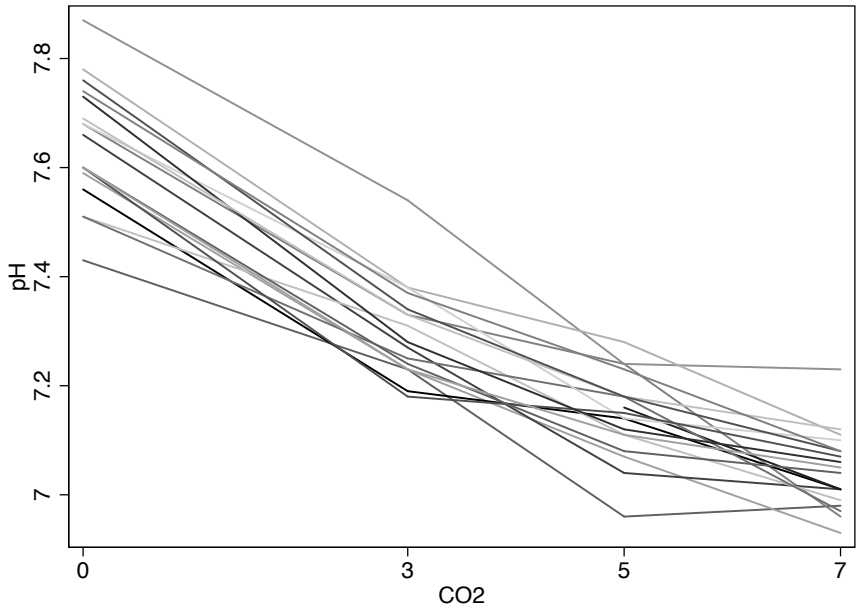
7.2.3 Hierarchical Models

Thus far we only discussed repeated measures within a single sampled unit, typically the person. However, one can generalize such a design to any number of nested units. For instance, common designs in educational research are to sample children, within classrooms, within schools,... Thus, one could consider sources of unmeasured variability that arise because children within classrooms are different, classrooms within schools, etc. Thus, a hierarchy of units leads to a hierarchy of random variables, and to so-called hierarchical mixed models. There is a rich literature on estimating parameters according to such models (e.g., see [**?**, **?**]). We will present a simple example from the world of opthmalogical research below, but it serves only as a gateway to a potentially rich class of models, much used in many social sciences.

Example: eyes within visits within people

Consider a functional opthalmologic study of the impact of $C0_2$ levels on pH in eyes [**?**]. Eighteen subjects, at two different visits, were fit with specialized goggles that regulated exposure to CO_2 (different levels for each eye) and subsequent measures were made of pH. Thus, the data consist of 18x4=72 nested measurements of eyes within visits within subjects. Thus, consider the outcome, pH, to be Y_{ijk}, for the ith person, jth visit, kth eye: $i = (1, \ldots, m = 18), j = (1, 2), k = (1, 2)$, where k is an arbitrary ordering of the eyes within a visit. Thus, for each of 18 subjects, there are 2 eyes measured at each of two visits, for a total of four measurements. The four experimentally manipulated values of CO_2 are

Figure 7.1: Line plot of pH versus CO_2



randomized among these four measures (see Figure 7.1).

As we emphasized in chapter 2, such plots are useful in thinking about how to model the observed variability, and this chapter puts such modeling into action. For instance, we can see that there is differences among the subjects in their pH at $CO_2 = 0$. Thus, we might include a random intercept at the individual level. What is not so clear from this plot is how much of this variability is due to differences in pH due to the visit. That becomes hard to distinguish graphically, since we do not get to observe a subject at different visits, but with the same CO_2 levels. Thus, models become a compelling way to estimate such sources of variability. However, buyer beware, when is difficult or impossible to explore sources of variability by empirically examining that variability across units at fixed levels of other

factors, it generally implies that our inferences will be based upon (sometimes strong) modeling assumptions. Thus, these models open up avenues for estimation unavailable by simpler, more intuitive approaches, but there is typically no free lunch - they only do so because such models entail stronger, often unverifiable, assumptions. We discuss these opportunities and pitfalls in more detail later in the chapter.

Consider a sequence of mixed effects models

$$\begin{aligned}
Y_{ijk} &= \beta_0 + \beta_1 X_{ijk} + \beta_2 X_{ijk}^2 + e_{ijk} \text{ (Model 1)} \\
Y_{ijk} &= \beta_0 + \beta_{0i} + \beta_1 X_{ijk} + \beta_2 X_{ijk}^2 + e_{ijk} \text{ (Model 2)} \\
Y_{ijk} &= \beta_0 + \beta_{0i} + \beta_{0ij} + \beta_1 X_{ijk} + \beta_2 X_{ijk}^2 + e_{ijk} \text{ (Model 3)} \\
Y_{ijk} &= \beta_0 + \beta_{0i} + \beta_{0ij} + (\beta_1 + \beta_{1i}) X_{ijk} + \beta_2 X_{ijk}^2 + e_{ijk} \text{ (Model 4)},
\end{aligned}$$

where, as usual, the random effect/error terms all assumed to be mean 0, normal random variables, and:

- $\sigma_e^2 \equiv var(e_{ijk})$
- $\sigma_{\beta_{0I}}^2 \equiv var(\beta_{0i})$
- $\sigma_{\beta_{0IJ}}^2 \equiv var(\beta_{0ij})$
- $\sigma_{\beta_{1I}}^2 \equiv var(\beta_{1i})$

We will also assume independence of random effects at different hierarchical levels, but

allow for general dependence of random effects at same level, i.e., we allow for correlation among the $(\beta_{0i}, \beta_{1i}, \beta_{2i})$.

The results are in table 7.4. We include here estimates of all parameters returned directly by procedure, as well as a measure of association comparing the mean of outcome at two different values of CO_2. Note, in addition to the estimates of coefficient and variance terms, we also include a model-fit statistic, so the so-called Aikake Information Criteria, which we discuss in a moment. Given these are quadratic models, and thus the coefficients typically do not provide an easily interpreted measure of association, we also include a comparison of the mean at different levels of CO_2, specifically $E(Y_{ijk}|X_{ijk} = 5) - E(Y_{ijk}|X_{ijk} = 0)$, which is equal to $5 * \beta_1 + 25 * \beta_2$. First, if we take each of the models as the true model, we can examine what each implies about the variability of pH among subjects. Model 1 assumes that all observations are independent, that they all have equal variance, and that all subjects have the same relationship of pH to CO_2. Model 2, by including a random intercept at the subject level allows that the interprets are different for each subject, but the coefficients on CO_2 are the same. As we learned thus far, this model thus assumes that observations (the $Y_{ijk} \mid X_{ijk}$) covary within a subject, and any two observations (any j, k vs. j', k') on the same subject have equal covariance, which does not vary by subject, and as discussed above is equal to $\sigma^2_{\beta_{0I}}$.

(UNFINSHED BELOW)

Table 7.4: Estimates(SE) and Akaike's Information Criterion (AIC) of four hierarchical linear mixed effects models for pH versus $C0_2$

	Model 1		Model 2		Model 3		Model 4	
Parameter	Est	SE	Est	SE	Est	SE	Est	SE
β_0	7.6	.021	7.6	.021	7.6	.020	7.6	.025
β_1	-.14	.014	-.14	.008	-.14	.0078	-.14	.0074
β_2	.0073	.0019	.0074	.0011	.0073	.0011	.0074	.0010
$E(Y_{ijk}\|X_{ijk}=5) - E(Y_{ijk}\|X_{ijk}=0)$	-.50	.027	-.50	.016	-.50	.015	-.50	.016
$\sigma_{\beta_{0i}}$			.072	.014	.068	.014	.094	.019
$\sigma_{\beta_{0ij}}$					.021	.013		
$\sigma_{\beta_{1i}}$							.0064	.0025
σ_e	.088	.0075	.051	.0051	.047	.0057	.047	.0046
AIC	-109		-143		-170		-175	

Table 7.5: Comparisons of Naive and Robust Inference for Model 4

	Est	SE Naive	SE Robust
β_0	7.6	.025	.026
β_1	-.14	.0074	.0080
β_2	.0074	.0010	.0010
$E(Y_{ijk}\|X_{ijk}=5) - E(Y_{ijk}\|X_{ijk}=0)$	-.50	.016	.017
$\sigma_{\beta_{0i}}$	.094	.019	.017
$\sigma_{\beta_{1i}}$	.0064	.0025	.0014
σ_e	.047	.0046	.0039

7.3 Binary Outcomes

Returning to the data on the relationship of drug/alcohol use on teenage sexual activity, consider the following mixed effects logistic regression model

$$\text{logit}\left[Pr(Y_{ij} = 1|\alpha_i, X_{ij} = x_{ij}\right] = a + \alpha_i + bx_{ij} \tag{7.7}$$

where the random effects, α_i, are assumed to be independently drawn from a Gaussian distribution with mean 0 and variance τ^2. In addition to assuming that observations for different teenagers are independent, this model also indicates that, *given* α_i, repeated observations on the same (i^{th}) teenager are also independent. Saying this another way, any correlation of observations on the same teenager arises entirely from the assumption that teenagers possess individual innate tendencies (i.e. probabilities) for sexual activity—modulo other covariates including drug/alcohol use (X_{ij})—that vary from person to person. In this way, the correlation structure for repeated observations here most closely resembles the working exchangeable correlation structure of a marginal model.

In maximum likelihood estimation, the three unknown parameters are thus a, b, associated with the fixed effects, and τ^2, the variance of the random effects. Note that here, b measures the log odds ratio associated with drug/alcohol use *for a specific teenager*, and will be larger than the (marginal) log odds ratio, as discussed in Section X.X. Table X.X displays the maximum likelihood estimates along with their estimated standard deviations.

Table 7.6: Estimates Based on a Simple Mixed Effects Logistic Regression Model Relating Drug and Alcohol Use to Teenage Sexual Activity

Model	Parameter	Estimate	SD
$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \alpha_i + bx_{ij}$	a	-1.0785	0.1355
	b	0.3697	0.1579

Table 7.7: Estimates Based on a Mixed Effects Logistic Regression Model Relating Drug and Alcohol Use to Teenage Sexual Activity and Weekday of Report

Model	Parameter	Estimate	SD
$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = a + \alpha_i + bx_{ij} + c_1(Tues) + \cdots + c_6(Sun)$	a	-1.3147	0.2183
	b	0.3476	0.1600
	c_1	0.2909	0.2428
	c_2	0.2805	0.2393
	c_3	0.1406	0.2381
	c_4	-0.0036	0.2463
	c_5	0.3252	0.2404
	c_6	0.5947	0.2359

The estimate of τ is 1.3154, which leads to an estimated (exchangeable) correlation of 0.34 (how is this calculated?).

As discussed in Section X.X, the conditional maximum likelihood provides an alternate approach to estimating b in a mixed effects model like (X¿X) that only includes random

Table 7.8: Conditional Maximum Likelihood Estimates Based on a Simple Mixed Effects Logistic Regression Model Relating Drug and Alcohol Use to Teenage Sexual Activity

Model	Parameter	Estimate	SD
$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \alpha_i + bx_{ij}$	b	0.2800	0.1631

Table 7.9: Conditional Maximum Likelihood Estimates Based on a Mixed Effects Logistic Regression Model Relating Drug and Alcohol Use to Teenage Sexual Activity and Weekday of Report

Model	Parameter	Estimate	SD
$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = a + \alpha_i + bx_{ij} + c_1(Tues) + \cdots + c_6(Sun)$	b	0.2559	0.1661
	c_1	0.2828	0.2450
	c_2	0.2767	0.2424
	c_3	0.1523	0.2402
	c_4	-0.0212	0.2450
	c_5	0.3286	0.2430
	c_6	0.5660	0.2385

intercepts. Here, the conditional maximum likelihood estimate of b is x.xxxx. Only 86 teenagers are used in this analysis. Discuss

7.3.1 Conditional Likelihood Techniques

7.3.2 More than Two Levels

Good example is Bangladesh family study (slides 70-78; must get data)

7.4 GEE versus Mixed Models