CPH 931 — Fall 2008 — Dr. Charnigo

Lecture 4

Linear mixed modeling

Problem #1. Last week we discussed how to perform a repeated measures analysis of variance (ANOVA) on data from the model

$$Y_{ij} = \mu_j + \alpha_i + \epsilon_{ij}. \tag{1}$$

The fixed effects $\mu_1, \ldots, \mu_k$ are the expected responses in k strata determined by a single categorical explanatory variable, the random effects $\alpha_1, \ldots, \alpha_n$ reflect the n subjects' individual tendencies to score higher or lower than expected, and the error terms $\epsilon_{11}, \ldots, \epsilon_{nk}$ describe how the observed responses differ from the expected responses adjusted for the subjects' individual tendencies.

Although the error terms $\epsilon_{11}, \ldots, \epsilon_{nk}$ are independent, the repeated measurements $Y_{i1}, \ldots, Y_{ik}$ on subject i are allowed to be correlated because they share the random effect α_i. Even so, model (1) is limited because it accommodates only a single categorical explanatory variable. Today we will address the problem of accommodating multiple explanatory variables.

Problem #2. We began this semester looking at the linear regression model

$$Y_i = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_m x_{m,i} + \epsilon_i, \tag{2}$$

which can accommodate multiple explanatory variables that are any mix of continuous and dichotomous. (As noted in Lecture 2 of CPH 930, any categorical variable can be recoded into dichotomous variables.)

Unfortunately, we cannot assume that the errors $\epsilon_1, \ldots, \epsilon_n$ in (2) are independent when there are repeated measurements on the same subject. That the errors $\epsilon_1, \ldots, \epsilon_n$ are correlated implies the same for the measurements $Y_1, \ldots, Y_n$. Thus, we cannot use ordinary least squares to fit model (2) when there are repeated measurements on the same subject.

While methodology ("generalized estimating equations") exists for fitting model (2) when there are correlated errors, today we will pursue a different approach to address the problem of correlated measurements. Rather than changing the model-fitting procedure, we will alter the model itself.

Formulating a linear mixed model. Let Y_{ij} denote the j^{th} measurement of the response variable on subject i. Let $x_{1,ij}$ through $x_{m,ij}$ denote the corresponding values on m explanatory variables that are any mix of continuous and dichotomous. The linear mixed model is

$$Y_{ij} = \beta_0 + \beta_1 x_{1,ij} + \cdots + \beta_m x_{m,ij} + \alpha_i + \epsilon_{ij}. \tag{3}$$

We refer to the regression-like coefficients $\beta_0, \ldots, \beta_m$ as fixed effects. They describe the relationship between the expected response and the explanatory variables. We refer to $\alpha_1, \ldots, \alpha_n$ as random effects and to $\epsilon_{11}, \epsilon_{12}, \ldots, \epsilon_{nk}$ as error terms. They have the same interpretations as in model (1), and we make the same assumptions about them that we did last week: the random effects are normally distributed with mean 0 and variance σ^2_α; the error terms are normally distributed with mean 0 and variance σ^2_ϵ; and, all of these quantities are mutually independent.

Apart from a minor notational difference (e.g., Y_i versus Y_{ij}), model (2) is a special case of model (3) with $\alpha_1 = \cdots = \alpha_n = 0$. What may be less obvious is that model (1) is also a special case of model (3). To see this:

- fix β_0 at 0;
- set m equal to k;
- let $x_{1,ij}$ equal 1 if the j^{th} measurement on subject i has treatment 1

(let $x_{1,ij}$ equal 0 otherwise);

- equate β_1 to μ_1;
- let $x_{2,ij}$ equal 1 if the j^{th} measurement on subject i has treatment 2;
- equate β_2 to μ_2;
- ...
- let $x_{k,ij}$ equal 1 if the j^{th} measurement on subject i has treatment k;
- equate β_k to μ_k.

Thus, model (3) is both an extension of model (1) that addresses Problem #1 and an extension of model (2) that addresses Problem #2.

Remarks on linear mixed modeling. Several items are worth noting.

1. The fixed effects are interpreted like the coefficients in a linear regression model. For instance, β_1 is the change in the expected response corresponding to a one-unit increase in X_1 when $X_2, \ldots, X_m$ are fixed. (If X_1 is dichotomous and coded as 1 for "present" and 0 for "absent", then β_1 is the difference between the expected responses for someone in whom X_1 is present and for an otherwise similar person in whom X_1 is absent.)

2. The fixed effects $\beta_0, \ldots, \beta_m$ and the "variance components" σ^2_α, σ^2_ϵ can be estimated by maximum likelihood (ML). Other estimation methods, especially "restricted maximum likelihood" (REML), are sometimes employed. However, in today's example we will use ML.

3. Although today's example will consider data in which the same number of measurements is made on each subject ("balanced data"), a linear mixed model can be fit to data in which different numbers of measurements are made on different subjects ("unbalanced data").

4. Another interpretation for the random effect α_i is that, along with β_0, it defines a "personal intercept" of $\beta_0 + \alpha_i$ for subject i. With this interpretation, β_0 is the average of the personal intercepts.

5. Model (3) can be generalized by allowing each subject to have a "personal slope" for X_1, a personal slope for X_2, and so forth, in much the same way that each subject has a personal intercept. This is how we can allow for the possibility that the blocking variable may interact with the explanatory variables. However, having noted this possibility, we will not pursue it further in CPH 931.

Example of a linear mixed model. Last week we applied repeated measures ANOVA to the digit symbol substitution test (DSST) data in {DSST.xls}. Specifically, we assessed whether the expectation for CORRECTTRIALS varied by experimental CONDITION when we confined attention to TIME = 3 and REPLICAT = 0.

This assessment, while useful for illustrating repeated measures ANOVA, was rather limited from a scientific perspective. There was no real reason to confine attention to TIME = 3 and REPLICAT = 0, nor was there a reason to ignore potentially relevant explanatory variables such as the subject's sensation seeking status as judged by a personality test (SSSTAT = 1 for high sensation seeker, SSSTAT = 0 for low sensation seeker) and the subject's baseline performance on the DSST (CORRECTTRIALS_B).

Today we will fit a linear mixed model to the DSST data with $m = 10$, $n = 16$ (number of subjects), and $k = 24$ (number of observations per subject, not counting baseline measurements). Let us define:

- Y = CORRECTTRIALS;
- $X_1 = 1$ when DRUG = 0 (placebo), $X_1 = 0$ otherwise;
- $X_2 = 1$ when ACTIVITY = 0 (low sensation seeking activity);

- $X_3 = 1$ when TIME = 2;
- $X_4 = 1$ when TIME = 3;
- $X_5 = 1$ when SSSTAT = 0 (low sensation seeking personality);
- $X_6 = 1$ when DRUG = 0 and SSSTAT = 0;
- $X_7 = 1$ when ACTIVITY = 0 and SSSTAT = 0;
- $X_8 = 1$ when SSSTAT = 0 and TIME = 2;
- $X_9 = 1$ when SSSTAT = 0 and TIME = 3;
- X_{10} = CORRECTTRIALS_B.

Thus, we will be modeling the expectation for CORRECTTRIALS as a function of DRUG, ACTIVITY, TIME, SSSTAT, and CORRECTTRIALS_B. The terms with X_6 through X_9 will allow for the possibility that SSSTAT may interact with DRUG, ACTIVITY, and TIME.

Discussion questions. Today the discussion questions better serve our needs here than at the end of the lecture.

1. Consider someone who has a high sensation seeking personality and who scored 60 on the DSST at baseline. As a function of β_0 through β_{10}, what is the expected DSST score for such a person while on amphetamine following a high sensation seeking activity at TIME = 3?

2. What is the expected DSST score for such a person while on placebo following a low sensation seeking activity at TIME = 3?

3. What null hypothesis could we test to see whether SSSTAT should be removed from the model altogether?

SAS output, covariance parameter estimates. The SAS output for our linear mixed model is shown in {DSSTMixed.rtf}. The "Covariance Parameter Estimates" box on page 2 shows estimates of σ_α^2 and σ_ϵ^2.

We have $\hat{\sigma}_\alpha^2 = 7.6426$ and $\hat{\sigma}_\epsilon^2 = 14.6236$. The p-values pertain to Wald tests of null hypotheses that $\sigma_\alpha^2 = 0$ and that $\sigma_\epsilon^2 = 0$, respectively. (Recall from CPH 930 that a Wald test assesses the quotient of a parameter estimate divided by its standard error in reference to a standard normal distribution or, equivalently, the square of such a quotient in reference a chi-square distribution on one degree of freedom.) At the usual 0.05 significance level, or even at the 0.01 significance level, both null hypotheses would be rejected.

Some caution is required, however. Because these hypothesis tests are one-sided (we cannot have $\sigma_\alpha^2 < 0$ or $\sigma_\epsilon^2 < 0$), the usual rule for computing p-values from Z statistics does not work. Here we must use the relation $p = 1 - \Phi(Z)$ instead of $p = 2(1 - \Phi|Z|)$, where Φ denotes the standard normal cumulative distribution function appearing in the Z table of any introductory statistics textbook. Thus, as illustrated in the SAS output, the p-value for testing $\sigma_\alpha^2 = 0$ is $0.0082 = 1 - \Phi(2.40)$ rather than 0.0164.

SAS output, fit statistics. The "Fit Statistics" box on page 2 displays negative twice the log likelihood and the values of three model selection criteria.

We may use negative twice the log likelihood to implement a likelihood ratio test involving the fixed effects. For instance, suppose that we wanted to test the null hypothesis that $\beta_1 = \beta_6 = 0$. We could refit the model without X_1 and X_6, which would yield a new (larger) value for negative twice the log likelihood. If the new value exceeded 2161.6 by more than $\chi^2_{2,1-\alpha}$, then we would reject the null hypothesis that $\beta_1 = \beta_6 = 0$ at the α significance level. In general, the degrees of freedom for the chi-square reference distribution would equal the number of fixed effects set to zero under the null hypothesis.

The AIC, AICC, and BIC are model selection criteria that can be used to judge between competing models if a hypothesis test cannot resolve the matter. For instance, if we were considering one model that omitted DRUG altogether and another model that omitted ACTIVITY altogether, then no hypothesis test would help us to choose between these models because neither model could be simplified to the other by setting fixed effects equal to zero. (Statisticians would say that neither model is "nested" within the other.) As the SAS output helpfully indicates, lower values for the AIC, AICC, and BIC are preferred.

SAS output, solution for fixed effects. The "Solution for Fixed Effects" box on pages 2 and 3 supplies the estimates of β_0 through β_{10}.

We have $\widehat{\beta}_0 = 32.2292$, $\widehat{\beta}_1 = -1.2192$, $\widehat{\beta}_2 = 0.4034$, and so forth. The T statistics and p-values pertain to tests of null hypotheses that $\beta_0 = 0$, that $\beta_1 = 0$, that $\beta_2 = 0$, and so forth. These tests are based on "normal theory", which is to say that they directly use the assumptions of normally distributed random effects and error terms rather than appealing to the kind of large-sample theory invoked by the likelihood ratio test.

SAS output, tests of fixed effects. The "Test of Fixed Effects" box on page 3 presents F statistics and p-values (based on normal theory) that may be of interest to an investigator.

Notice that the test for DRUG has only one numerator degree of freedom, so this is not a test of whether DRUG should be removed from the model altogether. Rather, the test for DRUG examines $\beta_1 + 0.5\beta_6$, which is the average of β_1 (negative of the amphetamine effect for a high sensation seeker) and $\beta_1 + \beta_6$ (negative of the amphetamine effect for a low sensation seeker). Hence, using the language of two-way ANOVA from your introductory statistics course, we might refer to the test for DRUG as a test of "main effects".

Likewise, the tests for ACTIVITY, TIME, and SSSTAT are tests of main effects. The test for TIME has two numerator degrees of freedom, one less than the number of levels for TIME. (The baseline measurement at TIME = 1 is being treated as an explanatory variable rather than as an instance of the response variable.)

The test for DRUG*SSSTAT is a test for interaction between DRUG and SSSTAT. Since the amphetamine effect for a high sensation seeker is the same as that for a low sensation seeker if and only if $\beta_6 = 0$, this test for interaction must be a test of $\beta_6 = 0$. In fact, the SAS output confirms this, as the p-value of 0.0143 in the "Test of Fixed Effects" box is identical to the corresponding p-value of 0.0143 in the "Solution for Fixed Effects" box.

Likewise, the tests for ACTIVITY*SSSTAT and SSSTAT*TIME are tests for interaction. The numerator degrees of freedom equals the number of levels for the first factor less one multiplied by the number of levels for the second factor less one.

Finally, the test for CORRECTTRIALS_B repeats the previous test of $\beta_{10} = 0$ from the "Solution for Fixed Effects" box.

SAS output, estimates. The "Estimates" box on page 4 presents estimates of user-specified linear combinations of fixed effects.

In the first row we find that the estimated expected response is 66.52 for the person specified in today's first discussion question. The T statistic and p-value are based on normal theory and test the (ridiculous) null hypothesis that the expected response for such a person is 0.

In the second row we find that the estimated expected response is 65.70 for the person specified in today's second discussion question.

The third row looks at the difference in expected responses between the people specified in today's first and second discussion questions. The estimate is 0.82, which makes sense since $66.52 - 65.70 = 0.82$, but we can use the T statistic and p-value to test the (not ridiculous) null hypothesis that the expected response for a high sensation seeker at TIME = 3 does not differ from the {amphetamine, high sensation seeking activity} experimental condition to the {placebo, low sensation seeking activity} experimental condition.

The fourth row is like the third row but pertains to a low sensation seeker. Interestingly, the {amphetamine, high sensation seeking activity} experimental condition is distinguished from the {placebo, low sensation seeking activity} experimental condition at TIME = 3 among low sensation seekers (p-value < 0.0001) even though no such distinction was made for high sensation seekers (p-value $= 0.2967$).

SAS output, contrasts. The "Contrasts" box on page 4 presents results for user-specified hypothesis tests based on normal theory.

In this example I requested only one hypothesis test. The null hypothesis was that from today's third discussion question. Note that the test has five numerator degrees of freedom, which makes sense because the null hypothesis sets five of the fixed effects equal to zero.