Topics for today:

- *Inference for random effects in an LMM*
- *Inference for variance components in an LMM (briefly)*

Associated reading: Sections 4 and 5 of 'LMM: inference' course notes), Verbeke (with a focus on Ch. 7), Hedeker (Chapters 4-7)

4 Estimation and tests for random effects (**b**)

- Although we can use ML or REML to estimate variance components, we may be interested in subject-specific random effect estimates.
  - In particular, they may allow us to determine if there are subjects with unusual trends relative to the rest of the group.
  - These subject-specific estimates cannot be derived from the marginal model.
  - A common approach is to use empirical Bayes (EB) estimators. EB estimators have an intuitive appeal since estimates are obtained essentially by taking a weighted average of personal and group-level data.

- *Example 1*: batting averages of Major League Baseball players.
  - At the beginning of the season, averages tend to vary more wildly (between 0.000 and 1.000).
  - As more games are played, the averages tend to settle into range between 0.200 and 0.350.
  - An EB estimate for a particular player near the beginning of the season may use a higher weight for the 'all-player' average and a lower average for that particular player to estimate that player's true average; later in the season the average may be weighted more heavily towards that player's particular average.

- *Example 2*: prevalence of a disease or illness for individual counties in a state.

    - Ideally, the best estimate of prevalence in a county would involve just the county data.

    - However, if collected data is sparse, then it might help to also base the estimate on state data as well.

    - The higher the variability in county data, the more the estimate is based on the state data, while the lower the variability in the county data, the more it is based on county data.

4.1 Empirical Bayes (EB) estimators for random effects

- In the Bayesian literature, the marginal distribution of **b** is called the prior distribution of the parameters **b** since it does not depend on the data **Y**. Once observed values of **Y** are obtained (**y**), the posterior distribution of **b**, which is $f(\mathbf{b}|\mathbf{y})$, can be calculated. Considering $\mathbf{b}_i$ and $\mathbf{Y}_i$ as the random effects and outcome data for individual i, the posterior distribution is

$$f(\mathbf{b}_i|\mathbf{Y}_i = \mathbf{y}_i) = \frac{f(\mathbf{y}_i|\mathbf{b}_i)f(\mathbf{b}_i)}{\int f(\mathbf{y}_i|\mathbf{b}_i)f(\mathbf{b}_i)d\mathbf{b}_i}$$

- In the expression above, the dependence of the density function on certain components of $\boldsymbol{\theta}$ is suppressed for notational convenience. The mean of this posterior distribution is a Bayes estimator of $\mathbf{b}_i$:

$$\begin{aligned} \hat{\mathbf{b}}_i(\boldsymbol{\theta}) \quad &= E(\mathbf{b}_i|\mathbf{Y}_i = \mathbf{y}_i) \\ &= \int \mathbf{b}_i f(\mathbf{b}_i|\mathbf{y}_i)d\mathbf{b}_i) \\ &= \mathbf{G}\mathbf{Z}_i^t\mathbf{V}_i^{-1}(\boldsymbol{\alpha})(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) \qquad (6) \end{aligned}$$

- The EB estimator is then computed by replacing unknown parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ with their ML or REML estimates (and hence the word 'empirical').

- We'll let $\hat{\mathbf{b}}_i\left(\hat{\boldsymbol{\theta}}\right) = \hat{\mathbf{b}}_i$ denote the Empirical Bayes estimator. For more detail, see section 7.2 in Verbeke.

- In terms of final notation, $\mathbf{b}$ and $\mathbf{Y}$ represent the vector of random effects and data, respectively, for the complete data, where data for subjects are stacked, while $\mathbf{b}_i$ and $\mathbf{Y}_i$ are the data for individual i.

4.2 The EB estimators and shrinkage

- Predicted values based on EB estimators for $\mathbf{b}_i$ are a weighted average of subject-specific data and group-averaged data, giving it an intuitive appeal:

$$
\begin{aligned}
\hat{\mathbf{Y}}_i &= \mathbf{X}_i\hat{\boldsymbol{\beta}} + \mathbf{Z}_i\hat{\mathbf{b}}_i \\
&= \mathbf{X}_i\hat{\boldsymbol{\beta}} + \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i^t\mathbf{V}_i^{-1}(\mathbf{Y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}) \\
&= (\mathbf{I}_{r_i} - \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i^t\mathbf{V}_i^{-1})\mathbf{X}_i\hat{\boldsymbol{\beta}} + \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i^t\mathbf{V}_i^{-1}\mathbf{Y}_i \\
&= \mathbf{R}_i\mathbf{V}_i^{-1}\mathbf{X}_i\hat{\boldsymbol{\beta}} + (\mathbf{I}_{r_i} - \mathbf{R}_i\mathbf{V}_i^{-1})\mathbf{Y}_i
\end{aligned}
$$

This is a weighted average of the estimated population average profile and the observed data.

- This demonstrates that $\hat{\mathbf{Y}}_i$ are shrunk towards the mean (relative to $\mathbf{Y}_i$).

- When residual variability (modeled through $\mathbf{R}_i$) is large in relation to between-subject variability (accounted for in $\mathbf{V}_i^{-1}$), the population-averaged profile ($\mathbf{X}_i\hat{\boldsymbol{\beta}}$) will have more weight, which makes sense since there is less certainty about individual data. (You can think of $\mathbf{R}_i$ as the "numerator" and $\mathbf{V}_i$ as the "denominator" in the quantity $\mathbf{R}_i\mathbf{V}_i^{-1}$.)

- Alternatively, when residual (within-subject) variability tends to be smaller and between-subject variability greater, then $\mathbf{Y}_i$ will have more weight.

- The EB estimators themselves exhibit the shrinkage property: $Var\left(\mathbf{L}\hat{\mathbf{b}}_i\right) \le Var\left(\mathbf{L}\mathbf{b}_i\right)$ for any $1 \times q$ real-valued vector $\mathbf{L}$. Remember also that $E\left(\mathbf{b}_i\right) = 0$. Thus the EB estimators are shrunk towards 0. For more detail, see Verbeke.

4.3 Inference associated with EB estimators

- The quantity $Var(\hat{\mathbf{b}}_i(\boldsymbol{\theta}))$ can be derived easily by substituting the MLE in for $\boldsymbol{\beta}$ and noting that it is a linear form of $\mathbf{y}_i$. (Laird and Ware, 1982, consider the Bayes estimator as in (6), but with $\boldsymbol{\beta}$ replaced with its MLE; they then derive theoretical results when covariance parameters are known or unknown.) The result is:

$$Var(\hat{\mathbf{b}}_i(\boldsymbol{\theta})) = \mathbf{G}_i \mathbf{Z}_i^t \{\mathbf{V}_i^{-1} - \mathbf{V}_i^{-1}\mathbf{X}_i \left(\sum \mathbf{X}_i^t \mathbf{V}_i^{-1} \mathbf{X}_i\right)^{-1} \mathbf{X}_i^t \mathbf{V}_i^{-1}\} \mathbf{Z}_i \mathbf{G}_i$$

- A few notes on this formula.

  - $Var(\hat{\mathbf{b}}_i(\boldsymbol{\theta}))$ is not the same as $Var(\mathbf{b}_i|\mathbf{Y}_i = \mathbf{y}_i)$; it is $Var[E(\mathbf{b}_i(\boldsymbol{\theta}) \mid \mathbf{y}_i)]$.

  - Second, for inference, $Var(\hat{\mathbf{b}}_i(\boldsymbol{\theta}) - \mathbf{b}_i)$ is used rather than $Var(\hat{\mathbf{b}}_i(\boldsymbol{\theta}))$ because the former take into account the variability in $\mathbf{b}_i$. This quantity is

$$Var(\hat{\mathbf{b}}_i(\boldsymbol{\theta}) - \mathbf{b}_i) = \mathbf{G}_i - Var(\hat{\mathbf{b}}_i(\boldsymbol{\theta}))$$

$$= \mathbf{G}_i - \mathbf{G}_i\mathbf{Z}_i^t\{\mathbf{V}_i^{-1} - \mathbf{V}_i^{-1}\mathbf{X}_i\left(\sum \mathbf{X}_i^t\mathbf{V}_i^{-1}\mathbf{X}_i\right)^{-1}\mathbf{X}_i^t\mathbf{V}_i^{-1}\}\mathbf{Z}_i\mathbf{G}_i$$

- In order to estimate $Var(\hat{\mathbf{b}}_i(\boldsymbol{\theta}) - \mathbf{b}_i)$ we typically just ‘plug in’ numerical values for unknown $\boldsymbol{\theta}$, not accounting for the added variability due to use of estimated values. In light of this, the selection of DF can help control the accuracy of inferential results for random effects, similar to that described previously for inference of fixed effects.

- t-tests can be constructed for random effects using relevant approximate t quantities.

    - For example, if $\mathbf{b}_i$ contains just a random intercept (i.e., $\mathbf{b}_i = b_{0i}$) then we can use $t = [(\hat{b}_{0i} - b_{0i}) - E(\hat{b}_{0i} - b_{0i})] / \hat{SE}(\hat{b}_{0i} - b_{0i})$, which reduces to $t = \hat{b}_{0i} / \hat{SE}(\hat{b}_{0i} - b_{0i})$ under the null, for the test of H_0: $b_{0i} = 0$.

    - For models with multiple random effect terms, we can carry out t-tests separately for each component of $\mathbf{b}_i$ (and subject). As before, the DF ($\hat{\nu}$) is ideally chosen to get the correct distribution of the test statistic under H_0; available methods to do this are as previously described.
    - Theory also exists for tests H_0: $\mathbf{Lb} = 0$ versus H_1: $\mathbf{Lb} \neq 0$. However, in practice, I have not yet found the need to use this.

- A 100(1–α)% confidence interval for an element b_{hi} of $\mathbf{b}_i$, is

$$\hat{b}_{hi} \pm t_{\hat{\nu},\alpha/2} \hat{SE}(\hat{b}_{hi} - b_{hi}).$$

- In SAS, when you request a solution for the random effects, the 'Estimate' will be numerical versions of (6), while 'Std Err Pred' is the square root of (diagonal elements of) $Var(\hat{\mathbf{b}}_i(\boldsymbol{\theta}) - \mathbf{b}_i)$. The calculated variance of the random effect estimates (using the 'population' version) will be the same as $V\hat{a}r(\hat{\mathbf{b}}_i(\boldsymbol{\theta}))$ (here, the hat on '*Var*' indicates that estimated values of $\boldsymbol{\theta}$ are 'plugged into' the calculation) and will be somewhat less than σ_b^2, reflecting the shrinking of the estimates back to the estimated population mean.

4.4 Computation of estimates and associated variances for random effects – see course notes

4.5 Empirical Bayes estimators for LMMs with random intercepts

- We have discussed Empirical Bayes estimators of random effects in mixed models. They have an intuitive appeal because they can be expressed as weighted averages of subject-specific information and population-average information.
  - The greater the variability of the subject data, the higher the weight is placed on the population average;
  - The more consistent the subject data is, the higher the weight is placed on the subject portion.
  - In previous notes, the weighted average was expressed for predicted values ($\hat{\mathbf{Y}}_i$) from an LMM.
  - It was briefly mentioned that the random effects estimates themselves ($\hat{\mathbf{b}}_i$) are shrunk towards the population mean (relative to $\mathbf{b}_i$), such that $Var\left(\hat{\mathbf{b}}_i\right) \leq Var\left(\mathbf{b}_i\right)$.
  - The amount of shrinkage depends on residual variance relative to subject variance. To study this further, we'll consider LMMs with random intercept terms.

- It was mentioned that the random effects estimates ($\hat{\mathbf{b}}_i$) are shrunk towards the population mean (relative to $\mathbf{b}_i$), such that $Var\left(\mathbf{L}\hat{\mathbf{b}}_i\right) \leq Var\left(\mathbf{L}\mathbf{b}_i\right)$ for a $1{\times}q$ real-valued vector $\mathbf{L}$.
    - A special case of this is $Var\left(\hat{b}_{hi}\right) \leq Var\left(b_{hi}\right)$, for h=1,…,q. This is easy to prove, since $Var(\hat{\mathbf{b}}_i - \mathbf{b}_i) + Var(\hat{\mathbf{b}}_i) = \mathbf{G}_i$, and the diagonal elements must be nonnegative.
    - The only time equality holds, such that $Var\left(\hat{b}_{hi}\right) = Var\left(b_{hi}\right)$, is when $Var(\hat{b}_{hi} - b_{hi}) = 0$. The amount of shrinkage in estimators depends on residual variance relative to subject variance. To study this further, we'll consider LMMs with random intercept terms.

- If the only random term in the model is an intercept term (for subjects) and $\mathbf{R}_i = \sigma^2\mathbf{I}$, (6) will reduce, since $\mathbf{G}$ only has one element (the variance of the random intercepts, call it σ_b^2), and $\mathbf{Z}_i^t$ is a row vector of 1's, call it $\mathbf{J}_{1\times r}$. For this case,

$$\mathbf{V}_i^{-1} = (\sigma_b^2\mathbf{J}_{r_i\times r_i} - \sigma_\varepsilon^2\mathbf{I}_{r_i\times r_i})^{-1} = (\mathbf{I}_{r_i\times r_i} - \frac{\sigma_b^2}{\sigma_\varepsilon^2 + r_i\sigma_b^2}\mathbf{J}_{r_i\times r_i})/\sigma_\varepsilon^2 .$$

- Ultimately, the Bayes estimator reduces to

$$\hat{b}_i(\boldsymbol{\theta}) = \lambda[\bar{Y}_i - (1/r_i)\sum_j \mathbf{X}_{ij}^r\boldsymbol{\beta}] \qquad (7)$$

where $\bar{Y}_i$ is the mean response for subject i, $\mathbf{X}_{ij}^r$ is the j^{th} row of $\mathbf{X}_i$, and

$$\lambda = \frac{r_i\sigma_b^2}{\sigma_\varepsilon^2 + r_i\sigma_b^2} .$$

- Note that λ is between 0 and 1; it is the weight used in the averaging of subject-specific and population average statistics. (Note also that u is unbolded since it involves just one estimator.) Greater between-subject variability relative to within-subject variability will yield larger values of λ (just like the ICC), but so will increasing the number of repeated measures.

- For practice, show that (7) holds, starting with (6). (You can use the given result for $\mathbf{V}_i^{-1}$.) When there is only a random intercept term and fixed intercept in the model $[Y_{ij} = \beta_0 + b_i + \varepsilon_{ij};\ b_i \sim iid\ N(0, \sigma_b^2)$, independently of $\varepsilon_{ij} \sim iid\ N(0, \sigma_\varepsilon^2)$; call it the 'simple random intercept model'], (7) becomes

$$\hat{b}_i(\boldsymbol{\theta}) = \lambda[\overline{Y}_i - \beta_0] \quad . \tag{8}$$

- We can consider λ as the shrinkage factor. What is being shrunk is the difference between the estimate of the random intercept for subject *i* and the population mean. If we add the population mean, β_0, we get the estimate for subject *i* in context of the population:

$$\lambda[\overline{Y}_i - \beta_0] + \beta_0 = \lambda \overline{Y}_i + (1-\lambda)\beta_0, \qquad (9)$$

which is a weighted average of $\overline{Y}_i$ and β_0.

- In practice we typically replace unknown parameters λ (which involves σ_b^2 and σ_ε^2) and β_0 in (8) and (9) with their estimators, yielding EB estimators.

- For the simple random intercept model, the variance of the Bayes estimator is

$$Var[\hat{b}_i(\boldsymbol{\theta})] = \sigma_b^2 \lambda \left(\frac{n-1}{n} \right) \quad . \qquad (10)$$

Verify for practice.

- As noted earlier, the variance quantity normally used in inference to account for randomness in $\mathbf{b}_i$ is

$$Var[\hat{b}_i(\boldsymbol{\theta}) - b_i] = \sigma_b^2 - Var[\hat{b}_i(\boldsymbol{\theta})]$$
$$= \sigma_b^2 - \sigma_b^2 \lambda \left(\frac{n-1}{n} \right) = \sigma_b^2 (1-\lambda) \left(\frac{n-1}{n} \right) \tag{11}$$

- Since the variance of EB estimators is more difficult to tackle, we usually work with the variance quantities of the Bayes estimators, in (10) and (11). But in practice we do then typically plug in values of unknown variances in the quantity, which I will denote as $\hat{Var}[\hat{b}_i(\boldsymbol{\theta})]$ and $\hat{Var}[\hat{b}_i(\boldsymbol{\theta}) - b_i]$.

- For the random intercept model we know that $Var(\hat{b}_i(\boldsymbol{\theta}) - b_i) + Var(\hat{b}_i(\boldsymbol{\theta})) = \sigma_b^2$ (more generally, that $Var(\hat{\mathbf{b}}_i(\boldsymbol{\theta}) - \mathbf{b}_i) + Var(\hat{\mathbf{b}}_i(\boldsymbol{\theta})) = \mathbf{G}_i$). For fixed variances, we know that $\lambda \rightarrow 1$ as the number of repeated measures, r_i is increased (and also n), in which case $Var[\hat{b}_i(\boldsymbol{\theta})] \rightarrow \sigma_b^2$, and hence $Var[\hat{b}_i(\boldsymbol{\theta}) - b_i] \rightarrow 0$.

- Let's continue with the simple random intercept model to see how estimation works, and how the variances look. The table below summarizes values of estimates, considering a range of parameter values and number of repeated measures, for n=10. Here, we consider β_0=3 and $\bar{Y}_i$=3.5.

β_0	σ_b^2	σ_ε^2	r_i	$\bar{Y}_i$	λ	$\hat{b}_i(\boldsymbol{\theta})$	$\hat{b}_i(\boldsymbol{\theta})+\beta_0$
3	**1**	4	5	3.5	0.56	0.28	3.28
3	**2**	4	5	3.5	0.71	0.36	3.36
3	**4**	4	5	3.5	0.83	0.42	3.42
3	**6**	4	5	3.5	0.88	0.44	3.44
3	**8**	4	5	3.5	0.91	0.45	3.45
3	**1**	4	20	3.5	0.83	0.42	3.42
3	**2**	4	20	3.5	0.91	0.45	3.45
3	**4**	4	20	3.5	0.95	0.48	3.48
3	**6**	4	20	3.5	0.97	0.48	3.48
3	**8**	4	20	3.5	0.98	0.49	3.49
3	**1**	4	100	3.5	0.96	0.48	3.48
3	**2**	4	100	3.5	0.98	0.49	3.49
3	**4**	4	100	3.5	0.99	0.50	3.50
3	**6**	4	100	3.5	0.99	0.50	3.50
3	**8**	4	100	3.5	1.00	0.50	3.50

When subject data are stronger (i.e., larger r_i), the estimate is weighted more heavily towards the subject average. In addition, when between-subject variance is relatively large, the estimate is weighted more towards subject data even when r_i is somewhat lower.

- The next table focuses on variance quantities. This demonstrates how shrinkage of estimates changes with sample size. When both n and r_i are relatively large, the shrinkage is not as great (in these cases, subject data is strong so there is less need to incorporate population-averaged data).

				n=10			n=100		
σ_b^2	σ_ε^2	r_i	λ	$Var[\hat{b}_i(\boldsymbol{\theta})]$	$Var[\hat{b}_i(\boldsymbol{\theta})] / \sigma_b^2$	$Var[\hat{b}_i(\boldsymbol{\theta}) - b_i]$	$Var[\hat{b}_i(\boldsymbol{\theta})]$	$Var[\hat{b}_i(\boldsymbol{\theta})] / \sigma_b^2$	$Var[\hat{b}_i(\boldsymbol{\theta}) - b_i]$
1	4	5	0.56	0.50	50.0%	0.50	0.55	55.0%	0.45
2	4	5	0.71	1.29	64.3%	0.71	1.41	70.7%	0.59
4	4	5	0.83	3.00	75.0%	1.00	3.30	82.5%	0.70
6	4	5	0.88	4.76	79.4%	1.24	5.24	87.4%	0.76
8	4	5	0.91	6.55	81.8%	1.45	7.20	90.0%	0.80
1	4	20	0.83	0.75	75.0%	0.25	0.83	82.5%	0.18
2	4	20	0.91	1.64	81.8%	0.36	1.80	90.0%	0.20
4	4	20	0.95	3.43	85.7%	0.57	3.77	94.3%	0.23
6	4	20	0.97	5.23	87.1%	0.77	5.75	95.8%	0.25
8	4	20	0.98	7.02	87.8%	0.98	7.73	96.6%	0.27
1	4	100	0.96	0.87	86.5%	0.13	0.95	95.2%	0.05
2	4	100	0.98	1.76	88.2%	0.24	1.94	97.1%	0.06
4	4	100	0.99	3.56	89.1%	0.44	3.92	98.0%	0.08
6	4	100	0.99	5.36	89.4%	0.636	5.90	98.3%	0.10
8	4	100	1.00	7.16	89.6%	0.836	7.88	98.5%	0.12

- In the table above, we don't have variance of EB estimators. But the variance of the Bayes estimator is indeed less than the true variance; note that the difference gets smaller and smaller as r_i increases. The variance quantity we use in practice is $\hat{Var}[\hat{b}_i(\boldsymbol{\theta}) - b_i]$. What we would like to use is $Var[\hat{b}_i(\hat{\boldsymbol{\theta}}) - b_i]$.

5 *Tests for variance components*

- We can use 'COVTEST' as an option in the PROC MIXED statement for tests involving covariance parameters, using Wald Z tests.

- We can also test for a 'significant additions' of random terms to a model (e.g., when including the random slope to an LMM with a random intercept) using likelihood ratio test methods. Here we compare changes in $-2ln(L)$ between models, which has an asymptotic chi-square distribution with DF= difference in the number of covariance parameters between the 2 models.

- For both approaches, tests are more valid when certain regularity conditions hold. See Verbeke, pages 64-66 for more detail.