Topics for these notes:

- *Distributions associated with the mixed model*
- *General parameter estimation in the mixed model*
- *Inference for fixed effects*
- *Properties of estimators in a mixed model*
- *Impact of modeling correlation on inference for fixed effects*

Associated reading: Sections 1-3, 6 and 7 in 'LMM: inference' course notes.

1 Distributions associated with the mixed model

1.1 The conditional distribution of **Y** *given* **b**

- The conditional distribution of **Y** given the random effects **b** is

$$\mathbf{Y} \mid \mathbf{b} \sim \mathrm{N}\left[\mathbf{X}\beta + \mathbf{Z}\mathbf{b}, \mathbf{R}\right] = \mathrm{N}\left[\begin{pmatrix}\mathbf{X} & \mathbf{Z}\end{pmatrix}\begin{pmatrix}\beta \\ \mathbf{b}\end{pmatrix}, \mathbf{R}\right]$$

- Note that if there are no random effects (in **b**) and $\mathbf{R} = \sigma^2\mathbf{I}$, then the model reduces to a general linear model.

- The classical method to analyze longitudinal data, “RM ANOVA”, essentially makes inference using this conditional distribution, since the random effects are treated as fixed effects. Adjustments are then made to tests in order to make ‘correct’ inference for estimators that account for the clustered data. In some cases this approach may yield the same or similar results as fitting a linear mixed model, but generally is much more limited.

1.2 The marginal distribution of $\mathbf{Y}$

- The joint distribution of $\mathbf{Y}$ and $\mathbf{b}$ is

$$\begin{pmatrix} \mathbf{Y} \\ \mathbf{b} \end{pmatrix} \sim N\left[\begin{pmatrix} \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{Z}\mathbf{G}\mathbf{Z}^t + \mathbf{R} & \mathbf{Z}\mathbf{G} \\ \mathbf{G}^t\mathbf{Z}^t & \mathbf{G} \end{pmatrix} \right].$$

- The marginal distribution of $\mathbf{Y}$ can be obtained by integrating out the random effects $\mathbf{b}$ from the joint distribution to obtain

$$\mathbf{Y} \sim N\left[\mathbf{X}\boldsymbol{\beta}, \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}^t + \mathbf{R} \right] .$$

- Modern mixed model methodology maximizes the likelihood associated with **Y**, or it equivalently minimizes

$$\ell = -2ln(L) = r_{tot}\ ln\left(2\pi\right) + ln\left|\mathbf{V}\right| + \left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)^t \mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)$$

  in order to make inferences about the regression coefficients **β** and covariance parameters **α**.

- The likelihood is also often expressed as a combination of subject-specific components:

$$L\left(\theta\right) = \prod_{i=1}^{n} \{\left(2\pi\right)^{-r_i/2} \left|\mathbf{V}_i\left(\boldsymbol{\alpha}\right)\right|^{-1/2} e^{-\left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}\right)^t \mathbf{V}_i^{-1}\left(\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta}\right)/2}\}$$

$$\Rightarrow \ell = -2ln(L) = \sum_{i=1}^{n} r_i ln\left(2\pi\right) + \sum_{i=1}^{n} ln\left|\mathbf{V}_i\left(\boldsymbol{\alpha}\right)\right| + \sum_{i=1}^{n}\left(\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta}\right)^t \mathbf{V}_i^{-1}\left(\boldsymbol{\alpha}\right)\left(\mathbf{Y}_i - \mathbf{X}\boldsymbol{\beta}\right)$$

2 *General parameter estimation in the mixed model*

- For the standard GLM, there are the regression coefficients (**β**) and one covariance parameter (σ^2) to estimate, which can be carried out using matrix algebra.

- Due to the inclusion of more covariance parameters in the model (in either **G** or **R**), parameter estimation in the mixed model is not as straightforward and generally requires at least some numerical analysis.

- Before describing these techniques in more detail, we will first discuss the most common estimation approaches, *maximum likelihood (ML)* estimation and *restricted maximum likelihood (REML)* estimation. There is also the *MIVQUE0* estimation approach, which is seldom used.

2.1 Maximum Likelihood (ML) Estimation

- The ML estimators of $\boldsymbol{\beta}$ are obtained by maximizing the likelihood L or minimizing ℓ (both given on previous page) based on the marginal distribution of $\mathbf{Y}$. This can be accomplished by first noting that maximizing the likelihood with respect to $\boldsymbol{\beta}$, conditional on $\boldsymbol{\alpha}$, yields

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \left(\sum_{i=1}^{n} \mathbf{X}_i^t \mathbf{V}_i^{-1}(\boldsymbol{\alpha}) \mathbf{X}_i \right)^{-} \sum_{i=1}^{n} \mathbf{X}_i^t \mathbf{V}_i^{-1}(\boldsymbol{\alpha}) \mathbf{Y}_i \quad \text{(subject-specific form)}$$

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \left(\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X} \right)^{-} \mathbf{X}^t \mathbf{V}^{-1} \mathbf{Y} \quad \text{(complete-data form)} \qquad (1)$$

where $\mathbf{V}_i = \text{Var}(\mathbf{Y}_i) = \mathbf{Z}_i \mathbf{G}_i \mathbf{Z}_i^t + \mathbf{R}_i$ (subject-specific form).

- Notice that we need values of $\boldsymbol{\alpha}$ in order to solve (1). To accomplish this, we can replace $\boldsymbol{\beta}$ in the likelihood function with its MLE in (1). Now we have a likelihood expressed in terms of $\boldsymbol{\alpha}$ only. Such a likelihood is sometimes referred to as a *profile likelihood*.

- Now we can maximize the profile likelihood function in order to obtain $\hat{\boldsymbol{\alpha}}$ using a numerical technique such as a ridge-stabilized Newton-Raphson algorithm (common in SAS). We can then go back and determine $\hat{\boldsymbol{\beta}}$ using (1) by replacing $\boldsymbol{\alpha}$ with its ML estimates.

- The MLE solution we obtain with this approach is the same as what we would obtain if we were able to maximize the likelihood simultaneously with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Notice that the estimator of $\boldsymbol{\beta}$ in (1) is identical to the weighted least-squares estimates with $\mathbf{V}^{-1}$ as the weighting matrix.

- One drawback of ML estimation is that associated estimators of covariance parameters tend to be biased. REML offers one way to remove or reduce bias.

- Note that (1) uses a generalized inverse in case $\mathbf{X}$ does not have full rank. If $\mathbf{X}$ does have full rank, then we can replace $\left(\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X}\right)^-$ with $\left(\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X}\right)^{-1}$. Issues of model parameterization and estimation here are analogous to those discussed in the GLM review.

2.2 Restricted maximum likelihood (REML) estimation

- To first introduce REML estimation, consider estimating the population variance based on a random sample from the population of interest.

- We know the sample variance (s^2, which uses 'n–1' in the denominator) is unbiased for the population variance in the case that the population mean is unknown and estimated (i.e., the usual case).

- But the ML estimator of σ^2 has n in the denominator. This demonstrates that ML estimates may not necessarily be unbiased estimators. REML estimation offers an alternative to ML estimation which helps to circumvent this problem. [Note: some call s^2 the adjusted MLE estimator.]

2.2.1 REML estimation for σ^2

- Let $\mathbf{J}_n = \underset{n\times 1}{\mathbf{J}}$, $\mathbf{I}_n = \underset{n\times n}{\mathbf{I}}$, and let $\mathbf{Y} = (Y_1,\ Y_2,\ \ldots,\ Y_n)^t$, where $\mathbf{Y} \sim N\left(\mu\,\mathbf{J}_n,\ \sigma^2\mathbf{I}_n\right)$

- Let **A** = any matrix with n–1 independent columns orthogonal to $\mathbf{J}_n$. E.g.:

$$\mathbf{A}=\begin{pmatrix} 1 & 0 & \ldots & 0 \\ -1 & 1 & \ldots & \ldots \\ 0 & -1 & \ldots & \ldots \\ \ldots & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & 1 \\ 0 & 0 & & -1 \end{pmatrix}$$

- Let $\mathbf{U} = \mathbf{A}^t\mathbf{Y}$ be "error contrasts." Note that $\mathbf{U} \sim N\left(\mathbf{0},\ \sigma^2\mathbf{A}^t\mathbf{A}\right)$ and that σ^2 is the only parameter in the distribution for **U**. Maximizing the likelihood for **U** with respect to σ^2 yields: $\hat{\sigma}^2 = [\mathbf{Y}^t\mathbf{A}(\mathbf{A}^t\mathbf{A})^{-1}\mathbf{A}^t\mathbf{Y}]/(n-1) = s^2$.

- In a similar fashion, it can be shown that the REML estimator of σ^2 in a GLM is $[1/(n-k)]\mathbf{Y}^t\left(\mathbf{I}-\mathbf{P}_\mathbf{X}\right)\mathbf{Y}$. Can you do this?

2.2.2 REML estimation in the linear mixed model

- Let **A** be a full rank matrix with columns orthogonal to the columns of **X**. Then $\mathbf{U} = \mathbf{A}^t\mathbf{Y} \sim N\left(\mathbf{0},\ \mathbf{A}^t\mathbf{V}(\boldsymbol{\alpha})\mathbf{A}\right)$, which does not depend on **β**. The associated likelihood is

$$L = (2\pi)^{-(r_{tot}-k)/2} \left|\sum_{i=1}^{n} \mathbf{X}_i^t\mathbf{X}_i\right|^{1/2} \left|\sum_{i=1}^{n} \mathbf{X}_i^t\mathbf{V}_i^{-1}\mathbf{X}_i\right|^{-1/2} \prod_{i=1}^{n} |\mathbf{V}_i|^{-1/2}\, e^{-1/2\sum_{i=1}^{n}(\mathbf{Y}_i-\mathbf{X}_i\hat{\boldsymbol{\beta}})^t\mathbf{V}_i^{-1}(\mathbf{Y}_i-\mathbf{X}_i\hat{\boldsymbol{\beta}})},$$

  where $k=rank(\mathbf{X})$.

- Note that this restricted L does not involve **β** parameters ($\hat{\boldsymbol{\beta}}$ is a function of **α**, as are $\mathbf{V}_i$ matrices) and is not a profile likelihood, as before. This is why some software (e.g., SAS) does not penalize for β terms in the AIC.

- The restricted likelihood can be maximized to yield $\hat{\boldsymbol{\alpha}}$. The problem is that this method really only offers a way to estimate parameters in $\boldsymbol{\alpha}$, not $\boldsymbol{\beta}$.

- The common approach to estimate $\boldsymbol{\beta}$ is then to plug the REML estimators of $\boldsymbol{\alpha}$ back into equation (1). But equation (1) was derived using ML methods, so this estimation of $\boldsymbol{\beta}$ is really based on a hybrid of ML and REML methods. Specifically, estimators of $\boldsymbol{\beta}$ use the ML form, but employ REML estimators of the variance components in that form.

- Verbeke denotes these as "REML" estimators of $\boldsymbol{\beta}$ (quotes emphasized). Since estimation of $\boldsymbol{\beta}$ is not based on one clear method, some statisticians prefer ML estimation. On the other hand, this estimation method does offer a way to reduce bias in variance component estimators. Some might argue that this is more important than the methodological issue.

2.3 Choosing the estimation method in SAS

- In the PROC MIXED statement, an option can be added: method = ML <or> REML <or> MIVQUE0 (no slash to separate the method option from the rest of the statement) if ML estimates are of interest. Note that the default method is REML; if no option is specified, then REML will be used.

2.4 The rank of $\mathbf{X}$ *and calculation of* $\hat{\boldsymbol{\beta}}$

- As mentioned, equation (1) can handle less-than-full-rank $\mathbf{X}$, in which case SAS PROC MIXED uses a generalized inverse. Specifically, linearly dependent columns that are found while moving from left to right are 'dropped' (just like with PROC GLM).

- Let $\mathbf{X}_{red}$ denote the reduced matrix. The inverse of $\mathbf{X}_{red}^{t}\mathbf{V}^{-1}\mathbf{X}_{red}$ can then be computed, after which columns and rows of 0's are added back into the resulting matrix, corresponding to the columns that were dropped.

- For example, if you dropped the 5^{th} column in $\mathbf{X}$, then you would add 0's in the (new) 5^{th} column and 5^{th} row of $\left(\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X}\right)^-$. This results in highest levels of factors or interactions being set to 0. Equivalent results could be obtained if the particular set-to-0 restrictions were placed directly on the model to make $\mathbf{X}$ full rank.

- The MLE property of $\hat{\boldsymbol{\beta}}$ will be maintained even if $\mathbf{X}$ does not have full rank. However, we still need to be concerned with estimability of parameters. In particular, we need to consider row vectors $\mathbf{L}$ such that $\mathbf{L}\boldsymbol{\beta}$ is estimable.

2.5 Varying the parameters in the **R** *matrix between groups of subjects*

- Usually, the parameters in the **R** matrix are assumed to be the same for all subjects. However, it is possible to use the GROUP option in the REPEATED statement in SAS to allow different groups of subjects to have difference covariance parameter estimates. In other words, the structure can be the same, but the actual estimated values will be allowed to differ for groups specified by the GROUP option.

- I think this would make sense if there is a clear grouping variable that results in a few (maybe 2, 3 or 4) group of subjects.

- Be careful about adding parameters to the model – using too many groups will introduce many parameters into the model. Remember, the power with inference comes in averaging over a random sample…

- The numbers of groups you define may depend somewhat on the sample size – the more data you have, the more reasonable it is to add parameters to the model.

*3 Estimation and tests for regression coefficients (**β**)*

3.1 The distribution of $\hat{\boldsymbol{\beta}}$

- Under the marginal model, $\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \left(\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}^t\mathbf{V}^{-1}\mathbf{Y}$ is distributed normally with mean $\boldsymbol{\beta}$ and variance

$$Var(\hat{\boldsymbol{\beta}}) = \left(\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}^t\mathbf{V}^{-1}Var(\mathbf{Y})\mathbf{V}^{-1}\mathbf{X}\right)\left(\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X}\right)^{-1}, \qquad (2)$$

  using the 'complete data' form. Can you show (2) starting with (1)?

- Since $Var(\mathbf{Y}) = \mathbf{V} = \mathbf{ZGZ}^t + \mathbf{R}$ based on our model, this further simplifies to

$$Var(\hat{\boldsymbol{\beta}}) = \left(\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X}\right)^{-1} \qquad (3)$$

- Unknown values of $\boldsymbol{\alpha}$ in (3) can be replaced with ML or REML estimates, yielding $\hat{Var}(\hat{\boldsymbol{\beta}})$. This variance is useful for inference regarding $\boldsymbol{\beta}$.

- Since the variance quantities above depend on $\boldsymbol{\alpha}$, the specification of the correct covariance matrix is important.

- When $\mathbf{X}$ does not have full rank, then we can write (3) as $Var(\tilde{\boldsymbol{\beta}}) = \left(\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X}\right)^-$. Miss-specifying the covariance structure (e.g., using the compound symmetry structure when it is in fact something else) may lead to biased estimates of variance, and hence inaccurate estimation and test results.

- An alternative is to use a robust estimator of variance that employs expression (2), but where $Var(\mathbf{Y}_i)$ is replaced with $(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}})(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}})^t$ and unknown $\boldsymbol{\alpha}$ replaced with estimates. This robust (or empirical or sandwich) variance is consistent, as long as the mean part of the model is correctly specified.

- Empirical *SE*s can be obtained using the EMPIRICAL option in the PROC MIXED statement (model-based *SE*s based on (3) are the default).

- Later we will see that empirical *SE*s are the default for PROC GENMOD with GEE, which is used to model longitudinal data with binary or count outcomes.

3.2 Confidence intervals and hypothesis tests

- Tests involving parameters of **β** can be carried out using approximate Wald, *t* or *F* tests. The *t* and *F* tests are generally preferred since unlike the Wald test, they take into account variability introduced by estimating **α**.

3.2.1 Hypothesis tests and confidence intervals using t-distribution methodology

- Analogous to the GLM case, we can conduct inference for **Lβ** as long as it is estimable. For the following, we consider such estimable functions of parameters. Denote the estimator of **β** as $\tilde{\boldsymbol{\beta}}$ that may not be unique due to the fact that **X** does not have full rank. How does $SE(\mathbf{L}\tilde{\boldsymbol{\beta}})$ from GLM compare with the comparable model-based *SE* from LMM? Remember that for GLM, $SE(\mathbf{L}\tilde{\boldsymbol{\beta}}) = \sqrt{\sigma^2 \mathbf{L}(\mathbf{X}^t\mathbf{X})^{-}\mathbf{L}^t}$. We can then estimate this *SE* by replacing the unknown population variance with the sample variance.

- For the mixed model, the model-based standard error is:

$$SE(\mathbf{L}\tilde{\boldsymbol{\beta}}) = \sqrt{\mathbf{L}(\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{L}^t} \qquad (4)$$

- But this *SE* will reduce to what we have for GLM when there are no random effects other than error, and **R** has the 'independent' structure ($\mathbf{R} = \sigma^2\mathbf{I}$).

- In this case, $\mathbf{V}^{-1}$ reduces to $\left(1/\sigma^2\right)\mathbf{I}$, so that the *SE*'s are the same! Again, unknown $\boldsymbol{\alpha}$ parameters in **V** in (4) are usually replaced with estimates, resulting in the underestimation of variability.

- One way to help account for this in methods of inference is to select the 'proper' degrees of freedom for the distribution associated with the test statistic, discussed below. <u>For practice</u>: derive (4) using model-based variance of **Y**, i.e., $Var\left(\mathbf{Y}\right) = \mathbf{ZGZ}^t + \mathbf{R} = \mathbf{V}$. (Hint: recall the linear form results from the GLM notes.)

- The test for H_0: $\mathbf{L\beta}=0$ (versus $H_0{}^C$) can be carried out by considering the quantity $t=(\mathbf{L}\tilde{\boldsymbol{\beta}})/\hat{SE}(\mathbf{L}\tilde{\boldsymbol{\beta}})$ that has an approximate *t*-distribution.

- To get the best approximation, we need to estimate the DF; denote this estimate as $\hat{\nu}$. (This is not like the GLM case where we have a nice ANOVA table and clear DF to use; here we need to estimate the DF.

- But this can be a good thing since it may allow us to help account for the unaccounted variability in the SE due to the use of estimated $\boldsymbol{\alpha}$ parameters. See the next section for more information on methods to select the DF.)

- So we carry out the test by assuming our test statistic *t* has a *t*-distribution with $\hat{\nu}$ DF under H_0.

- The confidence interval for $\mathbf{L\beta}$ has the form:

$$\mathbf{L}\hat{\boldsymbol{\beta}} \pm t_{\hat{\nu},\alpha/2}\,\hat{SE}(\mathbf{L}\hat{\boldsymbol{\beta}}).$$

3.2.2 F-tests

- *F*-tests can be used for tests H_0: $\mathbf{C\beta} = 0$ (vs. H_0^C). The form of the *F*-statistic under H_0 is

$$F = \left[\hat{\boldsymbol{\beta}}^t \mathbf{C}^t [\mathbf{C}\left(\mathbf{X}^t \mathbf{V}^{-1}(\hat{\boldsymbol{\alpha}})\mathbf{X}\right)^{-1} \mathbf{C}^t]^{-1} \mathbf{C}\hat{\boldsymbol{\beta}}\right] / rank(\mathbf{C}) \qquad (5)$$

  where the numerator DF is rank(**C**) and the denominator DF (DDF) can be estimated from the data just as the DF is estimated for *t*-tests.

- The distribution of the *F*-statistic in (5) has an approximate $F_{rank(L),\hat{\nu}}$ distribution, where the DDF is estimated.

- *F*-tests associated with main effects and interactions are given in default SAS PROC MIXED output. *F*-tests associated with linear combinations of **β** can be obtained in SAS using the CONTRAST statement.

3.3 Estimating the (denominator) DF for tests involving $\boldsymbol{\beta}$

- *t* and *F*-tests are performed somewhat differently when using LMM methods compared with GLM. However, for some models the test results will be essentially the same or very similar.

- With GLM we use ANOVA methods which break total sums of squares (SS) into sources, and then use this information along with DF to construct *t* or *F*-tests of interest. (See course notes for detail on Type I and III SS.)

- However, the standard GLM tests assume that we have independent observations and are thus not likely to be accurate for longitudinal or clustered data unless we employ repeated measures ANOVA or MANOVA techniques (invoked by including RANDOM or REPEATED statements in PROC GLM).

- For LMMs, we have test statistics that are functions of model parameters. After using ML or REML to estimate the parameters, we can then compute a test statistic described above. In order to get the appropriate distribution of the test statistic to determine a p-value, we need to estimate the DDF; there are several methods to do this. Sometimes they yield the same DDF.

- The DDF method (DDFM) can be specified in the MODEL statement of PROC MIXED in order to select the DDF. (For *t*-tests, the quantity is just DF, for *F*-tests, it is DDF; but we just use DDF for simplicity.)

- One of the key issues is accurately estimating the true distribution of the test statistic under the null hypothesis. The choice of DDF affects the variance of this distribution (and hence the p-value for the test).

- Thus, we can feasibly account for the correct variance of $\hat{\boldsymbol{\beta}}$ even though that quantity is underestimated in the test statistic (when the unknown $\boldsymbol{\alpha}$ parameters in that variance quantity are substituted with real numbers).

- To summarize, tests for the GLM involve SS and DF quantities that are broken down by sources. In LMMs, we have approximate test statistics that are computed and we can gain accuracy for the distribution of these test statistics by selection of the DDF.

- However, for simpler models the GLM and LMM approaches may yield equivalent tests results or close to it.

- In SAS, the denominator degrees of freedom methods (DDFM) that can be specified in the MODEL statement are:
  - CONTAIN (containment – default with RANDOM statement)
  - BETWITHIN (between-within – default for REPEATED but no RANDOM statement)
  - RESIDUAL (comparable to using the DF for the MSE in standard GLM, but usually not recommended since it often overestimates the optimal DF)
  - SATTERTH (performs a general Satterthwaite approximation to the DDF)
  - KENWARDROGER (SAS suggests this method when there is a REPEATED statement with TYPE=UN, but no RANDOM statement).

- You can in fact select your own denominator degrees of freedom using DDF=<*value*> Thus, if you feel you have a method that gives you an even better approximation to the null distribution through specification of the denominator DF, you can specify it. For more details, see Verbeke and Molenberghs, Linear Mixed Models in Practice, Springer, 1997, Appendix A, and also the SAS Help Documentation.

3.4 Illustrating test differences between PROC GLM and PROC MIXED

- Consider again the Beta Carotene data. We had 23 subjects in 4 groups. Each subject was measured over time, with up to 5 repeated measures. There are a total of 115 observations.

- For the basic GLM that does not account for repeated measures, there are 20 model DF (1 for intercept, 3 for group, 4 for time, 12 for group*time; both group and time are modeled as class variables here). Thus, there are 95 residual DF.

- The table below illustrates differences in quantities used for *F*-tests. In all cases, the Numerator DF for Group, Time and Group*Time are 3, 4 and 12.

Model approach; DDFM	*Group*	*Time*	*Group*Time*
Standard GLM	DDF=95; F=6.49; p=0.0005	DDF=95; F=6.45; p=0.0001	DDF=95; F=0.37; p=0.97
*Mixed with random int.; containment, betwithin, or satterth	DDF=19; F=1.52; p=0.24	DDF=76; F=34.66; p<0.0001	DDF=76; F=1.99; p = 0.04
Mixed with random int.; residual	DDF=95; F=1.52; p=0.21	DDF=95; F=34.66; p<0.0001	DDF=95; F=1.99; p=0.03
Mixed with random int., AR(1) structure for **R**; containment or betwithin	DDF=19; F=1.55; p=0.23	DDF=76; F=33.72; p<0.0001	DDF=76; F=1.92; p=0.045
Mixed with random int., AR(1) structure for **R**; satterth	DDF=18.6; F=1.55; p=0.23	DDF=34.4; F=33.72; p<0.0001	DDF=34.4; F=1.92; p=0.07

*Will get the same results if using repeated measures ANOVA.

- Note that if the mixed model is the same (rows 2 and 3; rows 4 and 5), then F-statistic values will be the same within columns. The difference shows up in the p-values (albeit somewhat small here) due to the different DDFM (and hence DDF) used.

- Also note that standard GLM (row 1) is not close to Mixed with random intercept using RESIDUAL DDF (row 3) mainly because there is not a term for subject within group in the former. In particular, look at the *Group***Time* interaction significance between these models!

- This goes to show just how powerful a random intercept can be. The residual DDFM is usually not recommended for actual use; it is just included here for demonstration.

6 Properties of estimators in a mixed model (also see SAS Help and Documentation)

- If **G** and **R** are known,
  - $\hat{\boldsymbol{\beta}}$ is the *best linear unbiased estimator* (BLUE) of $\boldsymbol{\beta}$
  - $\hat{\mathbf{b}}(\boldsymbol{\theta})$ is the *best linear unbiased predictor* (BLUP) of **b**.

- Here, "best" means minimum mean squared error.

- However, **G** and **R** are usually unknown and are estimated using one of the aforementioned methods.
  - These estimates, $\hat{\mathbf{G}}$ and $\hat{\mathbf{R}}$, are therefore employed in the estimation.
  - In this case, the BLUE and BLUP acronyms no longer apply, but the word *empirical* is often added to indicate such an approximation.
  - The appropriate acronyms thus become EBLUE and EBLUP (e.g., $\hat{\mathbf{b}}\left(\hat{\boldsymbol{\theta}}\right) = \hat{\mathbf{b}}$ is EBLUP of **b**).

7 Impact of modeling correlation on inference for $\boldsymbol{\beta}$

- How does the choice for the covariance structure of $\mathbf{R}_i$ and inclusion of random effects and associated parameters affect inference for fixed effect ($\boldsymbol{\beta}$) parameters in an LMM?

- This is an important question, because if adding covariance parameters does make a difference on inference for the fixed effects, then one might argue that ignoring correlation and using a simpler model (even a standard linear model) may be sufficient if the primary interest is in inference for $\boldsymbol{\beta}$.

- In order to answer this question, we can consider how specifying the covariance structure (of $\mathbf{Y}$) affects (i) estimates of $\boldsymbol{\beta}$, (or $\hat{\boldsymbol{\beta}}$), (ii) estimates of variance of $\hat{\boldsymbol{\beta}}$ (or $\hat{Var}(\hat{\boldsymbol{\beta}})$) and (iii) tests involving $\boldsymbol{\beta}$.

- Estimates of $\boldsymbol{\beta}$ are usually not affected very much by specification of the covariance structure. However, the form of the maximum likelihood estimator of $\boldsymbol{\beta}$ does involve $\mathbf{V}$ and thus can alter the estimates; typically this impact will be relatively small.

- The variance of $\hat{\boldsymbol{\beta}}$ also involves $\mathbf{V}$. But unlike the small changes that usually occur in $\hat{\boldsymbol{\beta}}$ with different covariance structure specifications, individual elements of $\hat{Var}(\hat{\boldsymbol{\beta}})$ can be greatly affected.

- This is due not only to the fact that correlation between measures is taken into account, but also due to the way in which the correlation structure is specified.

- In order to illustrate, consider subjects with 4 repeated measures over time, where the only predictor in the model is time as a class variable. We may be interested in the mean response at each time point as well as differences in mean response between time points.

- Variances of mean responses at one time point may differ based on the type of covariance structure specified. For example, the simple, CS and AR(1) structures will necessarily each have the same variance across time points due to constraints imposed by their structures, while the UN structure allows different variances.

- Thus, if there are differences in variance over time, then these differences will not be picked up with the simple, CS and AR(1) structures since they are essentially averaged out.

- Consequently, inference at specific time points could differ quite a bit depending on the covariance structure selected. These differences in variance are due to how the covariance structure is specified, and not as much due to the fact that correlation is or isn't taken into account. On average, the variances over time are often about the same regardless of structure chosen.

- For comparisons of means between time points, there are further potential differences based on whether correlation is or isn't taken into account.

- Generally speaking, repeated measures within subjects over time are positively correlated. Taking this into account will reduce the variance of the difference in estimated means. [Recall $\text{Var}(X-Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X,Y)$.]

- There could be very large differences in variances that compare time points, between models that do and do not take the correlation into account; the bigger the correlation in responses, the bigger the difference.

- Most non-simple structures account for this positive correlation in some way (e.g., CS, AR(1), UN). The best structure should be chosen based not only on goodness-of-fit statistics such as AIC, but also what makes sense based on a priori reasoning. In practice I have not come across longitudinal data with significantly negatively correlated responses over time yet.

- Tests for the fixed effects will also depend on how the degrees of freedom are specified, which is dependent on how the covariance structure is specified.

- Recall in SAS that the method of estimating the (denominator) degrees of freedom used is based on whether there is a RANDOM or REPEATED statement in the code (or both).

- Most of the time the DF will not greatly impact the p-values as long as the correlation is taken into account in some fashion (i.e., if you have correlated data and you specify RANDOM and/or REPEATED statements).