THE ANOVA APPROACH TO THE ANALYSIS OF LINEAR MIXED EFFECTS MODELS

We begin with the following special case. Suppose $y_{ijk} = \mu + \tau_i + U_{ij} + e_{ijk}$

$(i = 1, \ldots, t; j = 1, \ldots, n; k = 1, \ldots, m)$

$\beta = (\mu, \tau_1, \ldots, \tau_t)'$, $U = (U_{11}, U_{12}, \ldots, U_{tn})'$

$e = (e_{11}, e_{12}, \ldots, e_{ttn})'$, $\beta \in \mathbb{R}^{t+1}$, an unknown parameter vector,

We can write the model as

$y = X\beta + Zu + \varepsilon$,

Where $X = \begin{bmatrix} 1 & 1 \otimes 1 \\ 1 & \frac{1}{tnx1} \otimes \frac{1}{tnx1} \end{bmatrix}$

and $Z = \frac{1}{tnxtn} \otimes \frac{1}{mx1}$

$\left[ \begin{array}{c} \mu \\ \varepsilon \end{array} \right] \sim N\left( \left[ \begin{array}{c} \mu \\ 0 \end{array} \right], \left[ \begin{array}{cc} \sigma^2_u I & 0 \\ 0 & \sigma^2_e I \end{array} \right] \right)$,

$\sigma^2_u, \sigma^2_e \in \mathbb{R}^+$, unknown variance components

This is the standard model for a CRD with $t$ treatments, $n$ experimental units per treatment, and $m$ observations per experimental unit.

\begin{align*}
\text{Source} & \quad \frac{df}{t-1} \\
\text{treatments} & \quad \frac{df}{t-1} \\
\text{exp. units (treatments)} & \quad t(n-1) \\
\text{obs. units (exp. units, treatments)} & \quad tn(m-1) \\
\text{c. total} & \quad tnm-1
\end{align*}
\[
\frac{\text{Sum of Squares}}{\text{Source}} = \begin{array}{c|c|c|c}
\text{df} & t-1 & t(n-1) & \frac{t(n-1)}{t-1} \\
\text{df} & \frac{t(n-1)}{t-1} & \frac{t(n-1)}{t-1} & \frac{t(n-1)}{t-1} \\
\text{tr+} & \frac{t(n-1)}{t-1} & \frac{t(n-1)}{t-1} & \frac{t(n-1)}{t-1} \\
\end{array}
\]

\[
\frac{\sum_{i=1}^{n} \sum_{k=1}^{m} (Y_{ik} - \bar{Y}_{..})^2}{\sum_{i=1}^{n} \sum_{k=1}^{m} (Y_{ik} - \bar{Y}_{..})^2}
\]

The expected mean squares are:

\[
E(MS_{tr+}) = \frac{\text{df} \times \frac{t(n-1)}{t-1}}{\text{df} \times \frac{t(n-1)}{t-1}} 
= \frac{\sum_{i=1}^{n} \sum_{k=1}^{m} (Y_{ik} - \bar{Y}_{..})^2}{\sum_{i=1}^{n} \sum_{k=1}^{m} (Y_{ik} - \bar{Y}_{..})^2}
\]

Now, \( \bar{Y}_{..}, \ldots, \bar{Y}_{..} \) are \( N(0, \sigma_0^2) \).

Thus, \( E(\sum_{i=1}^{n} (Y_{ik} - \bar{Y}_{..})^2) = (t-1) \sigma_0^2/n \).

Similarly, \( E(\sum_{i=1}^{n} (Y_{ik} - \bar{Y}_{..})^2) = (t-1) \sigma_0^2/n \).

It follows that

\[
E(MS_{tr+}) = \frac{\sum_{i=1}^{n} \sum_{k=1}^{m} (Y_{ik} - \bar{Y}_{..})^2}{\sum_{i=1}^{n} \sum_{k=1}^{m} (Y_{ik} - \bar{Y}_{..})^2}
\]

\[
E(MS_{tr+}) = \frac{\sum_{i=1}^{n} \sum_{k=1}^{m} (Y_{ik} - \bar{Y}_{..})^2}{\sum_{i=1}^{n} \sum_{k=1}^{m} (Y_{ik} - \bar{Y}_{..})^2}
\]

\[
E(MS_{tr+}) = \frac{\sum_{i=1}^{n} \sum_{k=1}^{m} (Y_{ik} - \bar{Y}_{..})^2}{\sum_{i=1}^{n} \sum_{k=1}^{m} (Y_{ik} - \bar{Y}_{..})^2}
\]

\[
E(MS_{tr+}) = \frac{\sum_{i=1}^{n} \sum_{k=1}^{m} (Y_{ik} - \bar{Y}_{..})^2}{\sum_{i=1}^{n} \sum_{k=1}^{m} (Y_{ik} - \bar{Y}_{..})^2}
\]

\[
E(MS_{tr+}) = \frac{\sum_{i=1}^{n} \sum_{k=1}^{m} (Y_{ik} - \bar{Y}_{..})^2}{\sum_{i=1}^{n} \sum_{k=1}^{m} (Y_{ik} - \bar{Y}_{..})^2}
\]
Similar calculations allow us to add the following column to our ANOVA table.

<table>
<thead>
<tr>
<th>Source</th>
<th>EMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>tr+</td>
<td>( \frac{\sigma_e^2}{\sigma_e^2 + m \sigma_u^2} + \frac{nm}{t-1} \sum_{i=1}^{t} (\tau_i - \bar{\tau})^2 )</td>
</tr>
<tr>
<td>xu(trt)</td>
<td>( \sigma_e^2 + m \sigma_u^2 )</td>
</tr>
<tr>
<td>ou(xu, trt)</td>
<td>( \sigma_e^2 )</td>
</tr>
</tbody>
</table>

The entire table could also be derived using matrices.

\[
X_1 = \begin{bmatrix} 1 \end{bmatrix}, \quad X_2 = \frac{I}{\text{tr}x} \otimes \frac{1}{nm-x}, \quad X_3 = \frac{I}{\text{tr}x \times n} \otimes \frac{1}{xnm}.
\]

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>tr+</td>
<td>( \frac{Y'(P_2 - P_1)Y}{\text{rank}(X_1) - \text{rank}(X_1)} )</td>
</tr>
<tr>
<td>xu(trt)</td>
<td>( \frac{Y'(P_3 - P_2)Y}{\text{rank}(X_2) - \text{rank}(X_3)} )</td>
</tr>
<tr>
<td>ou(xu, trt)</td>
<td>( \frac{Y'(I - P_3)Y}{tnm - \text{rank}(X_3)} )</td>
</tr>
<tr>
<td>c. total</td>
<td>( \frac{Y'(I - P_3)Y}{tnm - 1} )</td>
</tr>
</tbody>
</table>

Mean squares could be computed using

\[
E(Y'AY) = \text{tr}(A \Sigma) + [E(Y)]'A \Sigma E(Y),
\]

where \( \Sigma = \text{Var}(Y) = Z \Sigma Z' + \Theta \)

\[
\Theta = \sigma_u^2 \frac{I}{\text{tr}x} \otimes \frac{1}{xm} + \sigma_e^2 \frac{I}{xnm \times ntm}
\]

and \( E(Y) = \begin{bmatrix} \mu + \tau_1 \\ \mu + \tau_2 \\ \vdots \\ \mu + \tau_t \end{bmatrix} \otimes \frac{1}{xnm}. \)

Furthermore, it can be shown that

\[
\frac{Y'(P_2 - P_1)Y}{\sigma_e^2 + m \sigma_u^2} \overset{\text{t-df}}{\sim} \chi^2_{t-1} \left[ \frac{nm}{\sigma_e^2 + m \sigma_u^2} \sum_{i=1}^{t} (\tau_i - \bar{\tau})^2 / (t-1) \right]
\]

\[
\frac{Y'(P_3 - P_2)Y}{\sigma_e^2 + m \sigma_u^2} \overset{\text{t-df}}{\sim} \chi^2_{tnm-t}
\]

\[
\frac{Y'(I - P_3)Y}{\sigma_e^2} \overset{\text{t-df}}{\sim} \chi^2_{tnm-tn}
\]

and that these \( \chi^2 \) random variables are independent.
It follows that
\[ F_1^{c} = \frac{M_{S_{trt}}}{M_{S_{xu(trt)}}} \sim F \left[ \frac{nm}{\sigma_e^2 + m\sigma_u^2} \sum_{i=1}^{n} (\bar{z}_i - \bar{z})^2 / (t-1) \right] \]

\[ F_2^{c} = \frac{M_{S_{xu(trt)}}}{M_{S_{ou(xu, trt)}}} \sim \left( \frac{\sigma_e^2 + m\sigma_u^2}{\sigma_e^2} \right)^{F \left[ t_n - t, t_n - t \right]} \]

Thus, we can use \( F_1 \) to test \( H_0: \bar{z}_1 = \cdots = \bar{z}_t \) and \( F_2 \) to test \( H_0: \sigma_u^2 = 0 \).

Also note that
\[ E \left( \frac{M_{S_{xu(trt)}} - M_{S_{ou(xu, trt)}}}{m} \right) = \left( \frac{\sigma_e^2 + m\sigma_u^2}{\sigma_e^2} \right) - \frac{\sigma_e^2}{m} = \sigma_u^2. \]

Thus, \( \frac{M_{S_{xu(trt)}} - M_{S_{ou(xu, trt)}}}{m} \) is an unbiased estimator of \( \sigma_u^2 \).

This estimator of \( \sigma_u^2 \) is known as a "Method of Moments" estimator because it is obtained by equating observed statistics with their moments (expected values in this case) and solving the resulting set of equations for unknown parameters in terms of observed statistics.

Although
\[ \frac{M_{S_{xu(trt)}} - M_{S_{ou(xu, trt)}}}{m} \]

is unbiased for \( \sigma_u^2 \), it can take negative values even though \( \sigma_u^2 \), the variance of the \( u \) random effects, cannot be negative.
As we have seen previously,
\[ \operatorname{Var}(y) = \sigma_u^2 I_{tn \times tn} \otimes I_{m \times m} + \sigma_e^2 I_{tn \times tn} \]
\[ = \Sigma. \]
It turns out that
\[ \hat{\beta} = (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} y = (X' X)^{-1} X' y \]
Thus, the GLS estimator of any estimable \( \mathbf{C} \beta \) is equal to the OLS estimator in this special case.

If we define \( \xi_{ij} = u_{ij} + e_{ij} \), \( i, j \) and \( \sigma^2 = \sigma_u^2 + \sigma_e^2 / m \), we have
\[ \bar{y}_{ij} = \gamma + z_{i} + \bar{u}_{ij} \], where the \( \xi_{ij} \)'s are iid \( N(0, \sigma^2) \). Thus, averaging the same number \( (m) \) of multiple observations per experimental unit results in a Normal Theory Gauss-Markov linear model for the averages \( \bar{y}_{ij} : i = 1, \ldots, t; j = 1, \ldots, \frac{N}{m} \).

An Analysis Based on Averages for Each Experimental Unit:
Recall that our model is
\[ y_{ijk} = \mu + z_i + u_{ij} + e_{ijk} \quad (i = 1, \ldots, t; j = 1, \ldots, n_j; k = 1, \ldots, m) \]
The average of observations for experimental unit \( ij \) is
\[ \bar{y}_{ij} = \mu + z_i + u_{ij} + \bar{e}_{ij}. \]
Inferences about estimable functions of \( \beta \) obtained by analyzing these averages are identical to the results obtained using the ANOVA approach as long as the number of multiple observations per experimental unit is the same for all experimental units.
When using the averages as data, our estimate of $\sigma^2$ is an estimate of $\sigma_u^2 + \sigma_e^2/m$. We can't separately estimate $\sigma_u^2$ and $\sigma_e^2$, but this doesn't matter if our focus is on inference for estimable functions of $\beta$.

\[
\text{Var} \left( \bar{Y}_{i\cdot} \right) = \text{Var} \left( \frac{1}{n} \sum_{i=1}^{n} (u_{i\cdot} + e_{i\cdot}) \right) \\
= \text{Var} \left( \bar{u}_{i\cdot} + \bar{e}_{i\cdot} \right) \\
= \text{Var} \left( \bar{u}_{i\cdot} \right) + \text{Var} \left( \bar{e}_{i\cdot} \right) \\
= \frac{\sigma_u^2}{n} + \frac{\sigma_e^2}{nm} \\
= \frac{1}{n} \left( \sigma_u^2 + \sigma_e^2/m \right) \\
= \sigma^2/n.
\]

Because $E(Y) = \begin{bmatrix} m + \tau_1 \\ m + \tau_2 \\ \vdots \\ m + \tau_t \end{bmatrix} \otimes \frac{1}{nm}$, the only estimable quantities are linear combinations of the treatment means $m + \tau_1, m + \tau_2, \ldots, m + \tau_t$, whose Best Linear Unbiased Estimators are $\bar{Y}_{1\cdot}, \bar{Y}_{2\cdot}, \ldots, \bar{Y}_{t\cdot}$, respectively.

\[
\text{Var} \left( \bar{Y}_{i\cdot} \right) = \text{Var} \left( \frac{1}{n} \sum_{j=1}^{n} Y_{ij\cdot} \right) \\
= \frac{1}{n} \text{Var} \left( \bar{Y}_{i1\cdot} \right) \\
= \frac{1}{n} \text{Var} \left( \frac{1}{m} \sum_{m}^{1/m} (Y_{i1\cdot}, \ldots, Y_{i1m})' \right) \\
= \frac{1}{n} \frac{1}{m^2} \sum_{m}^{1/m} \left( \sigma_e^2 I + \sigma_u^2 \frac{1}{m} \right) \frac{1}{m} \\
= \frac{1}{n} \frac{1}{m^2} \left( \sigma_e^2 m + \sigma_u^2 m^2 \right) \\
= \frac{\sigma_e^2 + m \sigma_u^2}{nm} = \frac{\sigma^2}{n}.
\]
Thus, we don't need separate estimates of $\sigma_a^2$ and $\sigma_e^2$ to carry out inference for estimable $C$. We do need to estimate $\sigma^2 = \frac{\hat{\sigma}_e^2}{m}$. This can equivalently be estimated by $\frac{\text{MSE}}{m}$ or by the MSE in an analysis of the experimental unit means $\bar{y}_{i..}$. \[ \bar{y}_{i..} - \bar{y}_{j.} \]

A 100(1-\alpha)% confidence interval for $\bar{y}_{i..} - \bar{y}_{j.}$ is \[ \bar{y}_{i..} - \bar{y}_{j.} \pm t^{\alpha/2} \sqrt{\frac{2 \text{MSE}_{\text{error}}}{mn}} \]

A test of $H_0: \bar{y}_{1..} = \bar{y}_{2..}$ can be based on \[ t = \frac{\bar{y}_{1..} - \bar{y}_{2..}}{\sqrt{\frac{2 \text{MSE}_{\text{error}}}{mn}}} \]

Thus,

\[
\begin{align*}
\text{Var}(\bar{y}_{i..} - \bar{y}_{j.}) &= \text{Var}(\bar{y}_{i..}) + \text{Var}(\bar{y}_{j.}) \\
&= \frac{\sigma_a^2}{n} + \frac{\sigma_e^2}{m} \\
&= \frac{2 \sigma_e^2}{mn} \\
&= \frac{2 \text{MSE}_{\text{error}}}{mn}
\end{align*}
\]
What if the number of observations per experimental unit is not the same for all experimental units?

Let's look at two miniature examples.

\[ \begin{align*}
X_1 &= \frac{1}{2} , & X_2 &= X , & X_3 &= Z \\
MS_{\text{TRT}} &= \mathbf{y}'(P_2-P_1)\mathbf{y} = 2(\overline{y}_{11}-\overline{y}_{12})^2 + 2(\overline{y}_{21}-\overline{y}_{22})^2 \\
&= (\overline{y}_{11} - \overline{y}_{22})^2 \\
MS_{\text{XU(TRT)}} &= \mathbf{y}'(P_3-P_2)\mathbf{y} = (\overline{y}_{11} - \overline{y}_{12})^2 + (\overline{y}_{21} - \overline{y}_{22})^2 \\
&= \frac{1}{2} (\overline{y}_{11} - \overline{y}_{12})^2 \\
MS_{\text{OU(XU,TRT)}} &= \mathbf{y}'(\mathbf{I}-P_3)\mathbf{y} = (\overline{y}_{21} - \overline{y}_{22})^2 + (\overline{y}_{21} - \overline{y}_{22})^2 \\
&= \frac{1}{2} (\overline{y}_{21} - \overline{y}_{22})^2 \\
\end{align*} \]

\[
E(MS_{\text{TRT}}) = E((\overline{y}_{11} - \overline{y}_{22})^2) = E((\bar{\tau}_1 - \bar{\tau}_2 + \bar{u}_1 - \bar{u}_{21} + \bar{e}_{11} - \bar{e}_{21})^2) = (\bar{\tau}_1 - \bar{\tau}_2)^2 + \text{Var}(\bar{u}_1) + \text{Var}(\bar{u}_{21}) + \text{Var}(\bar{e}_{11}) + \text{Var}(\bar{e}_{21})
\]

\[
= (\bar{\tau}_1 - \bar{\tau}_2)^2 + \sigma_u^2/2 + \sigma_e^2/2 + \sigma_e^2/2 + \sigma_e^2/2
\]

\[
= (\bar{\tau}_1 - \bar{\tau}_2)^2 + 1.5 \sigma_u^2 + \sigma_e^2.
\]
\[ E(MS_{\text{wct/rep}}) = \frac{1}{2} E(Y_{21} - Y_{11})^2 = \frac{1}{2} (\frac{\sigma_a^2}{\sigma_e^2} + 2\sigma_e^2) = \frac{\sigma_a^2}{\sigma_e^2} + \sigma_e^2. \]

\[
E(MS_{\text{wct/rep}}) = \frac{1}{2} E(Y_{21} - Y_{11})^2 = \frac{1}{2} E(e_{21} - e_{11})^2 = \sigma_e^2.
\]

\[ F = \frac{MS_{\text{TRT}}}{\frac{MS_{\text{wct/rep}}}{\sigma_a^2 + \sigma_e^2}} \sim F_{1,1} \]

\[ F \sim \frac{1.5 \sigma_a^2 + \sigma_e^2}{0.5 \sigma_a^2 + \sigma_e^2} \]

\[ 1.5 \sigma_a^2 + \sigma_e^2 - 0.5 \sigma_a^2 = 1.5 \sigma_a^2 + \sigma_e^2. \]

The test statistic that we used to test

\[ H_0: \tau_1 = \ldots = \tau_k \text{ in the balanced case is not F distributed.} \]

\[ Case: \frac{MS_{\text{wct/rep}}}{MS_{\text{wct/rep}}} \]

But this is not a useable test statistic because it depends on unknown parameters.

We'd like our denominator to be an unbiased estimator of \( 1.5 \sigma_a^2 + \sigma_e^2 \) in this case.

Consider \( 1.5 MS_{\text{wct/rep}} - 0.5 \sigma_a^2 \). The expectation is 1.55, and 0.5 \( \sigma_a^2 \) = 0.5 \( \sigma_e^2 \).
The ratio
\[
\frac{MS_{\text{trt}}}{1.5MS_{\text{xu(trt)}} - 0.5MS_{\text{ox(xu,ter)}}}
\]
can be used as an approximate F statistic with 1 numerator d.f. and a denominator d.f. obtained using the Cochran-Satterthwaite method.

What does the BLUE of the treatment means look like in this case?
\[
\beta = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \quad X = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad Z = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}
\]

\[
\text{Var}(\bar{y}) = \Sigma = Z \Sigma Z' + R = \sigma^2_w Z Z' + \sigma^2_e I
\]

\[
= \sigma^2_w \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} + \sigma^2_e \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}
\]

The Cochran-Satterthwaite method will be explained in the next set of notes.

We shouldn't expect this approximate F-test to be reliable in this case because of our pitifully small dataset.

It follows that
\[
\hat{\beta} = (X'Z^{-1}X)^{-1}X'Z^{-1}Y
\]

\[
= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} Y
\]

Fortunately, this is a linear estimator that does not depend on unknown Variance components.
Consider a slightly different scenario:

\[
\hat{\beta} = (X'X)^{-1}X'Y
\]

\[
\sigma^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n-1}
\]

Of course, this is not an estimator because it is a function of unknown parameters. Replace \( \sigma^2 \) and \( \sigma^2 \) by estimates in the expression above.

\[
\hat{\beta} = \frac{2\sigma^2 + 2\sigma^2}{3\sigma^2 + 4\sigma^2} \left[ \begin{array}{c} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \end{array} \right]
\]

It is straightforward to show that the weights on \( Y_{11} \) and \( Y_{12} \) are:

\[
\frac{1}{\text{var}(Y_{11})} + \frac{1}{\text{var}(Y_{12})}
\]

and

\[
\frac{1}{\text{var}(Y_{11})} + \frac{1}{\text{var}(Y_{21})}
\]

respectively.

In this case, it can be shown that

\[
\hat{\beta} = (X'X)^{-1}X'Y
\]
... is an approximation to the BLUE.

\( \hat{\beta} \) is not even a linear estimator in this case.

Its exact distribution is unknown.

When sample sizes are large, it is reasonable to assume that the distribution of \( \hat{\beta} \) is approximately the same as the distribution of \( \hat{\beta} \).

**Summary of Main Points**

Many of the concepts we have seen by examining special cases hold in greater generality.

For many of the linear mixed models commonly used in practice, balanced data are nice because...

\[
\text{Var} \left( \hat{\beta} \right) = \text{Var} \left[ \left( X' \Sigma^{-1} X \right)^{-1} X' \Sigma^{-1} \mathbf{Y} \right] \\
= \left( X' \Sigma^{-1} X \right)^{-1} X' \Sigma^{-1} \text{Var} (\mathbf{Y}) \left( X' \Sigma^{-1} X \right)^{-1} \\
= \left( X' \Sigma^{-1} X \right)^{-1} X' \Sigma^{-1} X \left( X' \Sigma^{-1} X \right)^{-1} \\
= \left( X' \Sigma^{-1} X \right)^{-1}.
\]

\[
\text{Var} \left( \hat{\beta} \right) = \text{Var} \left[ \left( X' \Sigma^{-1} X \right)^{-1} X' \hat{\beta} \right] \\
= ? \sim \left( X' \Sigma^{-1} X \right)^{-1}.
\]

1. It is relatively easy to determine degrees of freedom, sums of squares, and expected mean squares in an ANOVA table.

2. Ratios of appropriate mean squares can be used to obtain exact F-tests.

3. For estimable \( C \hat{\beta} \), \( C \hat{\beta} \) is \( \hat{\beta} \).
   (The OLS estimator equals the GLS estimator.)
4. When \( \text{Var}(\hat{\xi}^2) = \text{constant} \times E(MS) \),

exact inferences about \( \hat{\xi}^2 \) can be

obtained by constructing \( t \) tests

or confidence intervals.

\[
t = \frac{\hat{\xi}^2 - \xi^2}{\sqrt{\text{constant} \times MS}} \sim t_{d.f.(MS)}
\]

5. Simple analysis based on experimental unit

averages gives the same results as those

obtained by linear mixed model analysis of

the full data set.

2. The estimator \( \hat{C^2} \) may be a nonlinear

estimator of \( C^2 \) whose exact distribution

is unknown.

3. Approximate inference for \( C^2 \) is often

obtained by using the distribution of

\( \hat{C^2} \), with unknowns in that distribution

replaced by estimates.

When data are unbalanced, the analysis

of linear mixed models may be considerably

more complicated.

1. Approximate F tests can be obtained

by forming linear combinations of

Mean Squares to obtain denominators

for test statistics.

Whether data are balanced or unbalanced,

unbiased estimators of variance

components can be obtained by the

method of moments.