Section 12.4 Mixed Models (comment)

Discussion of restricted model and unrestricted model. We will only consider the unrestricted model. This is what SAS uses. It refers to the assumptions for random effects...

All random effects are independent and are normally distributed with mean 0.

For example, \( Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha \beta)_ij + \varepsilon_{ijk} \)

\[ \alpha_i \sim N(0, \sigma^2_\alpha) \]
\[ \beta_j \sim N(0, \sigma^2_\beta) \]
\[ (\alpha \beta)_{ij} \sim N(0, \sigma^2_{\alpha \beta}) \]
\[ \varepsilon_{ijk} \sim N(0, \sigma^2) \]

Sections 12.6 & 12.7 from CLRT are replaced by the Lenth handout: "Simplified diagrams, models, and expected mean squares"
Recall the Lenth diagrams from Ch. 2 DDD.
We'll consider only 1 measurement per subject/cheese combo for now.

Example: cheese x background

SUBJECT

- Nested factors are located in a vertical chain.
  (Note: When choosing terms for your model, no interaction exists between factors in the same vertical chain)

- Crossed factors are listed in the horizontal direction.
  (This is also true for factors that appear in different locations, like cheese and SUBJECT above)

- Two factors that appear in different horizontal positions appear in "combination" (aka. crossed)

- When including superscripts to show the # of levels of a factor, the # of "actual" levels of a nested factor is larger than it first appears (this is why we may show 2 different numbers of levels for a nested factor).

\[ \text{cheese}^4 \times \text{background}^2 \]

10 levels of SUBJECT

20 "unique" subjects.

10 levels of SUBJECT

Within background.
Also:

- Nested factors are usually random.
- Fixed factors are usually crossed.

Example 1 from Lenth Handout:

Factors: fabric, temp, load

fixed fixed random

\begin{equation*}
\text{fabric, temp crossed,}
\end{equation*}

\begin{equation*}
\text{LOAD nested in temp.}
\end{equation*}

\begin{equation*}
\text{fabric crossed with LOAD.}
\end{equation*}

Lenth Simplified diagram:

\begin{equation*}
temp^3 \times \text{fabric}^4
\end{equation*}

\begin{equation*}
5\text{LOAD}^5
\end{equation*}

What terms go into the model?

Possible:

- temp
- fabric
- LOAD(temp)
- temp * fabric
- fabric * LOAD(temp)

Note: *show nesting now.

\begin{equation*}
temp + \text{LOAD(temp)}
\end{equation*}

- This will actually be our residual

\begin{equation*}
temp*LOAD(temp)
\end{equation*}

"doesn't exist" not legal
I'm doing based on "between-load" and "within-load" effects.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>temp</td>
<td>2</td>
</tr>
<tr>
<td>LOAD(temp)</td>
<td>12</td>
</tr>
<tr>
<td>fabric</td>
<td>3</td>
</tr>
<tr>
<td>temp * fabric</td>
<td>6</td>
</tr>
<tr>
<td>RESIDUAL (a.k.a. fabric * LOAD(temp))</td>
<td>36</td>
</tr>
<tr>
<td>C. TOTAL</td>
<td>59</td>
</tr>
</tbody>
</table>

Total 59 df.

Could also include d.f.
on diagram as subscript:

\[
\text{temp}^3 \times \text{fabric}^3
\]

\[
5 \text{LOAD}'15
\]

\[
4 \text{LOAD}'12
\]

Within each temp level, we use 4 d.f. forLOAD. So, we use 3x4=12 df for LOAD.
How do we get the EMS for our terms in the model? (unrestricted model)

Lenth's Rules for EMS (unrestricted model)

- Terms can be labeled
  - Fixed
  - Random
  - Mixed (but these are actually random)

0. Write each model term on a separate line.

1. Find each term’s EMS “leading component.”
   a) If term t is fixed, this is listed as Q(t).
      And Q(t) has a form which contains the
      summation of the squared fixed effects, like $\sum q_i^2$.
   b) If term t is random (or mixed),
      this is $\sigma_t^2$ multiplied by the # of times
      each level of the effect is repeated (not observed).

   For example, let t = LOAD (temp).
   "leading component" is $4 \sigma_{22}^2$
   for each load contains 4 observations in it.

   [For random terms that appear in the diagram, this
    "multiplier" is simply $(N/\text{Right superscript})$; for interaction
    terms it is $(N/(\text{multiplication of all relevant right superscripts}))].$

Recall:

\[
\begin{align*}
\text{LOAD (temp)} & \text{ is } 60/15 = 4 \\
\text{Fabric} \times \text{LOAD (temp)} & \text{ is } 60/(4 \times 15) = 1
\end{align*}
\]

Now we have all "leading components" specified.
Lenth Rules for EMS (unrestricted) cont.

2. For each term $t$, add to this leading component the leading component of all other random (or mixed) terms that contain every factor in $t$.

For example, let $t = \text{temp}$. Leading component is $Q(\text{temp})$.

"temp" is contained in...

\[
\text{LOAD(\text{temp})} \\
\text{temp} \times \text{fabric}
\]

fixed term, don't include

\[
\text{RESIDUAL} = \text{fabric} \times \text{LOAD(\text{temp})}
\]

So,

\[
E(\text{MS}_{\text{temp}}) = Q(\text{temp}) + 4 \sigma_\delta^2 + \sigma^2
\]

NOTE: RESIDUAL "contains" every term.

---

Example:

<table>
<thead>
<tr>
<th>temp</th>
<th>($Q(\text{temp})$) + 4 $\sigma_\delta^2$ + $\sigma^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOAD(temp)</td>
<td>($4 \sigma_\delta^2$) + $\sigma^2$</td>
</tr>
<tr>
<td>fabric</td>
<td>($Q(\text{fabric})$)</td>
</tr>
<tr>
<td>temp x fabric</td>
<td>($Q(\text{temp}\times\text{fabric})$)</td>
</tr>
<tr>
<td>RESIDUAL (a.k.a. fabric*LOAD(temp))</td>
<td>($\delta^2$)</td>
</tr>
</tbody>
</table>

From this table, we can choose appropriate numerators and denominators for F-tests of interest.
This particular example qualifies under a "split-plot" analysis. Each "top level" factor (temp) treatment (L, M, H) is randomly assigned to a load. Each load receives all levels of the "lower level" factor (fabric).

We have more resolution (less noise) for comparing fabrics as differences between fabrics are done within a load (and only contain 1 variance component $\sigma^2$).

We have less resolution (and fewer "effective" samples) for comparing temperatures as differences are done between loads (and still contain 2 variance components).

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>temp</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>LOAD(temp)</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>fabric</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>temp * fabric</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>RESIDUAL (or fabric*LOAD(temp))</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>C. TOTAL</td>
<td>59</td>
<td></td>
</tr>
</tbody>
</table>

Whole-plot analysis looks like a RBD as...

\[
F_0 = \frac{MS_{\text{temp}}}{MS_{\text{LOAD(temp)}}} \text{ or } \frac{MS_{\text{temp}}}{MS_{\text{LOAD(temp)}}} \overset{\text{H}_0}{\sim} F(2, 12)
\]

\[
F_0 = \frac{MS_{\text{fabric}}}{MSE} \overset{\text{H}_0}{\sim} F(3, 36)
\]
Model: \[ Y_{ijk} = \mu + \alpha^*_i + \zeta_j^*(\cdot) + S_k + (\Psi^*_i)_k + \varepsilon_{kj}(\cdot) \]

\[ \zeta_j^*(\cdot) \sim \text{iid } N(0, \sigma^2_{\zeta}(t)) \]
\[ \varepsilon_{kj}(\cdot) \sim \text{iid } N(0, \sigma^2) \]

Difference in 2 fabric means:
\[ \text{Var}(\overline{Y}_{..k} - \overline{Y}_{..m}) = \text{Var}(k + \overline{\alpha}^*_i + \overline{\zeta}_j(\cdot) + \overline{S}_k + (\overline{\Psi}^*_i)_k + \overline{\varepsilon}_{kj}(\cdot)) \]
\[ - (m + \overline{\alpha}^*_m + \overline{\zeta}_j(\cdot) + \overline{S}_m + (\overline{\Psi}^*_m)_m + \overline{\varepsilon}_{jm}(\cdot)) \]
\[ = \text{Var}(\overline{\varepsilon}_{k}(\cdot) - \overline{\varepsilon}_{m}(\cdot)) = \frac{\sigma^2}{15} + \frac{\sigma^2}{15} = \frac{2\sigma^2}{15} \]

We have 15 observations on each fabric.

Difference in 2 temp means:
\[ \text{Var}(\overline{Y}_{..j} - \overline{Y}_{..m}) = \text{Var}(j + \overline{\alpha}^*_i + \overline{\zeta}_j(\cdot) + \overline{S}_j + (\overline{\Psi}^*_j)_j + \overline{\varepsilon}_{..j}(\cdot)) \]
\[ - (m + \overline{\alpha}^*_m + \overline{\zeta}_j(\cdot) + \overline{S}_m + (\overline{\Psi}^*_m)_m + \overline{\varepsilon}_{..m}(\cdot)) \]
\[ = \text{Var}(\overline{\zeta}_j(\cdot) + \overline{\varepsilon}_{..j}(\cdot)) - (\overline{\zeta}_j(\cdot) + \overline{\varepsilon}_{..m}(\cdot)) \]
\[ = \frac{\sigma^2_{\zeta}(t)}{5} + \frac{\sigma^2}{20} + \frac{\sigma^2_{\zeta}(t)}{5} + \frac{\sigma^2}{20} = \frac{2[\sigma^2 + 4\sigma^2_{\zeta}(t)]}{20} \]

5 load observations in each temp \( j \)
20 observations in each temp \( t \)

(5 load obs. x 4 fabric observations)
If you want more power for detecting differences between temperatures, you'll want to do more loads (rather than putting 8 pieces of fabric in each of the given 15 loads). This will give you more d.f. for the error in the test for temp and it will reduce the effect of \( \sigma^2 \) in the contrast standard error (as a bonus, it will also reduce the effect of \( \sigma^2 \) in the standard error).

---

**Other examples**

**Pigment Example:** Rep nested in sample nested in batch.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>EMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>BATCH</td>
<td>14</td>
<td>( \frac{40^2}{\beta} + \frac{20^2}{5(\beta^2)} ) + ( \sigma^2 )</td>
</tr>
<tr>
<td>SAMPLE (BATCH)</td>
<td>15</td>
<td>( \frac{25^2}{5(\beta^2)} ) + ( \sigma^2 )</td>
</tr>
<tr>
<td>RESIDUAL</td>
<td>30</td>
<td>( \sigma^2 )</td>
</tr>
<tr>
<td>(a.k.a. REPSAMPLE (BATCH))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Total</td>
<td>59</td>
<td></td>
</tr>
</tbody>
</table>

*No interaction possible.*

\[ N = 60 \]
Penicillin Example (RCBD)

$\text{BATCH}^5 \times \text{process}^4$

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>EMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>BATCH</td>
<td>4</td>
<td>$10^2$ + $5^2$</td>
</tr>
<tr>
<td>process</td>
<td>3</td>
<td>$5^2$</td>
</tr>
<tr>
<td>RESIDUAL (aka. BATCH*process)</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>C. TOTAL</td>
<td>19</td>
<td></td>
</tr>
</tbody>
</table>

Terms:
- BATCH
- process
- BATCH*process = RESIDUAL
  - (no replication between the 2 factors)

$\text{EMS} = \frac{\text{EMS}}{\text{df}}$

Note: The d.f. for your terms in the ANOVA table should still add-up to the corrected total. If there are extra d.f. to be accounted for, include the RESIDUAL term and allocate these extra d.f. to this term.
New example: RCBD with Subsampling

Interested in studying 4 methods of tenderizing chicken. Three chickens are chosen at random, and the 4 methods are applied to "like" portions from each chicken. So, we are blocking on chicken or "animal." In actually reading tenderness each portion is broken into 5 parts, and a tenderness reading is taken on each part.

Sources of variability (or Factors):

Method, Animal, Reading (or subsampling)

"fixed" "random" "random"

\[ \text{ANIMAL}^3 \times \text{method}^4 \]

\[ \text{5 READING}^{60} \]

Possible terms: ANIMAL

method

ANIMAL \times \text{method}

READING(ANIMAL \times \text{method})

N = 60

\[ \frac{df}{2} \quad \frac{df}{3} \quad \frac{df}{6} \quad 48 \]

\[ \frac{\text{EMS}}{\frac{20\delta^2}{\sigma^2} + \frac{5\delta_{AM}^2}{\sigma^2} + \frac{5\delta^2}{\sigma^2}} \quad \frac{Q(m)}{5\delta_{AM}^2 + 5\delta^2} \quad \frac{\delta^2}{\delta^2} \]

Here, ANIMAL \times \text{method} is the error term for testing "method." The extra subsampling doesn't increase the "effective sample size for testing "method." (Want more power, get more animals.)
For \( Y_{ijk} = m + a_i + b_j + (a_i b_j) + \epsilon_{ijk} \) and the t.e. distributions which are independent.

The difference in 2 method means:

\[
\text{Var} (\bar{Y}_{i..} - \bar{Y}_{m..}) = \frac{\sigma_{Am}^2}{3} + \frac{\sigma^2}{15} + \frac{\sigma_{Am}^2}{3} + \frac{\sigma^2}{15} = \frac{2 [\sigma^2 + 5 \sigma_{Am}^2]}{15}
\]

To reduce the effect of \( \sigma_{Am}^2 \) on our standard error, we need more animals. If \( \sigma^2 \) was very large, then more subsampling would be useful, but it doesn't reduce the effect of \( \sigma_{Am}^2 \).

A 95% C.I. for \( (\bar{Y}_{i..} - \bar{Y}_{m..}) \):

\[
\bar{Y}_{i..} - \bar{Y}_{m..} \pm t_{0.025, df} \times \text{S.E. } (\bar{Y}_{i..} - \bar{Y}_{m..})
\]

\[
\text{S.E. } (\bar{Y}_{i..} - \bar{Y}_{m..}) = \sqrt{\frac{2 \cdot MS_{\text{ANIMAL*method}}}{15}}
\]

So, \( df = 6 \) since that is the df for \( \text{ANIMAL*method} \).
Returning to Lehn example 1...

What if we want to test the contrast of cell means?

\[ H_0: \text{fabric 1 low} = \text{fabric 1 medium} \]

\[
\text{Var} (\bar{Y}_{1.1} - \bar{Y}_{2.1}) = \text{Var} (\bar{Y}_{1.1}) + \text{Var} (\bar{Y}_{2.1}) = \frac{\sigma^2}{5} + \frac{\sigma^2}{5} = \frac{2\sigma^2}{5}
\]

Whether you do a t-I, or a statistical test, you'll need the df that goes with the standard error.

\[ \sigma^2 + \sigma_L^2 = \text{MSE} + \left( \frac{\text{MSE}_{L(c)} - \text{MSE}}{4} \right) = \frac{3}{4} \text{MSE} + \frac{1}{4} \text{MSE}_{L(c)} \]

\[ \text{S.E.}(\bar{Y}_{1.1} - \bar{Y}_{2.1}) = \sqrt{\frac{2}{5} \left( \frac{3}{4} \text{MSE} + \frac{1}{4} \text{MSE}_{L(c)} \right)} = \sqrt{\frac{3}{10} \text{MSE} + \frac{1}{10} \text{MSE}_{L(c)}} \]

and Satterthwaite \( df = \frac{\left( \frac{1}{10} \text{MSE}_{L(c)} + \frac{3}{10} \text{MSE} \right)^2}{\left( \frac{1}{10} \text{MSE}_{L(c)} \right)^2 + \left( \frac{3}{10} \text{MSE} \right)^2} \)
In general, when your variate of interest is a linear combination of mean squares...

\[
\hat{\mathbf{Y}} = a_1 MS_1 + a_2 MS_2 + \ldots + a_k MS_k
\]

then

\[
t_0 = \frac{\hat{Y}}{\sqrt{\text{Var}(\hat{Y})}} \quad \text{and} \quad t_0 \sim t_{df}
\]

with

\[
df = \frac{(a_1 MS_1 + a_2 MS_2 + \ldots + a_k MS_k)^2}{\frac{(a_1 MS_1)^2}{df_1} + \frac{(a_2 MS_2)^2}{df_2} + \ldots + \frac{(a_k MS_k)^2}{df_k}}
\]

and these are the Cochran-Satterthwaite degrees of freedom.