Nested (Hierarchical) Random Effects Model
Pigment Example
2 replicates from each of 2 samples from each of 15 barrels are measured.
The main goal of the analysis is estimation of variance components.

SAS statements for data input:

/* This is a SAS program for analyzing data from a nested or hierarchical experiment.

The data are measurements of moisture content of a pigment taken from Box, Hunter and Hunter (page 574). */

data set1;
  infile "Y:\pigment.dat";
  input batch sample rep y;
run;

proc print data=set1 (obs=10);
  run;

Obs  batch  sample  rep  y
    1      1      1      1  40
    2      1      1      2  39
    3      1      2      1  30
    4      1      2      2  30
    5      2      1      1  26
    6      2      1      2  28
    7      2      2      1  25
    8      2      2      2  26
    9      3      1      1  29
   10      3      1      2  28

SAS statements for getting expected mean squares:

/* The "random" statement in the following GLM procedure prints out formulas for expectations of mean squares. */

proc glm data=set1;
  class batch sample rep;
  model y = batch sample(batch);
  random batch sample(batch)/test;
  run;
The GLM Procedure

Class Level Information

<table>
<thead>
<tr>
<th>Class</th>
<th>Levels</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>batch</td>
<td>15</td>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13 14 15</td>
</tr>
<tr>
<td>sample</td>
<td>2</td>
<td>1 2</td>
</tr>
<tr>
<td>rep</td>
<td>2</td>
<td>1 2</td>
</tr>
</tbody>
</table>

Dependent Variable: y

<table>
<thead>
<tr>
<th>Sum of Source</th>
<th>DF</th>
<th>Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>29</td>
<td>2080.683333</td>
<td>71.747701</td>
<td>78.27</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>30</td>
<td>27.500000</td>
<td>0.916667</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>59</td>
<td>2108.183333</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Don’t look at the P-values in the following ANOVA tables, they are not correct. They all use MSE as the error term (Wrong). But the SS can be used to estimate variance components.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type I SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>batch</td>
<td>14</td>
<td>1210.933333</td>
<td>86.495238</td>
<td>94.36</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>sample(batch)</td>
<td>15</td>
<td>869.750000</td>
<td>57.983333</td>
<td>63.25</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>batch</td>
<td>14</td>
<td>1210.933333</td>
<td>86.495238</td>
<td>94.36</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>sample(batch)</td>
<td>15</td>
<td>869.750000</td>
<td>57.983333</td>
<td>63.25</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

The GLM Procedure

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Expected Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>batch</td>
<td>[ \text{Var}\left(\text{Error}\right) + 2 \text{Var}\left(\text{sample(batch)}\right) + 4 \text{Var}\left(\text{batch}\right) ]</td>
</tr>
<tr>
<td>sample(batch)</td>
<td>[ \text{Var}\left(\text{Error}\right) + 2 \text{Var}\left(\text{sample(batch)}\right) ]</td>
</tr>
</tbody>
</table>

Use the technique of setting the EMS equal to the observed MS to estimate the variance components. These are called the ANOVA estimates of the variance components. They are unbiased, but they can take negative values.

1. \[ E(MSE) = \sigma^2 \quad \Rightarrow \quad \hat{\sigma}^2 = MSE \quad \Rightarrow \quad \hat{\sigma}^2 = 0.9167 \]

2. \[ E(MS_{\text{Sample(Batch)}}) = \sigma^2 + 2\sigma^2_{\text{Sample(Batch)}} \]
   \[ \Rightarrow \quad \hat{\sigma}^2_{\text{Sample(Batch)}} = \frac{MS_{\text{Sample(Batch)}} - \hat{\sigma}^2}{2} = \frac{57.9833 - 0.9167}{2} = 28.5333 \]

3. \[ E(MS_{\text{Batch}}) = \sigma^2 + 2\sigma^2_{\text{Sample(Batch)}} + 4\sigma^2_{\text{Batch}} \]
   \[ \Rightarrow \quad \hat{\sigma}^2_{\text{Batch}} = \frac{MS_{\text{Batch}} - MS_{\text{Sample(Batch)}}}{4} = \frac{86.4952 - 57.9833}{4} = 7.1280 \]
If you include the **test** option in the **random** statement in PROC GLM, then you ask SAS to use appropriate error terms when conducting F-tests.

The GLM Procedure  
Tests of Hypotheses for Random Model Analysis of Variance

Dependent Variable: y

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>batch</td>
<td>14</td>
<td>1210.933333</td>
<td>86.495238</td>
<td>1.49</td>
<td>0.2256</td>
</tr>
<tr>
<td>Error</td>
<td>15</td>
<td>869.750000</td>
<td>57.983333</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error: MS(sample(batch))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>sample(batch)</td>
<td>15</td>
<td>869.750000</td>
<td>57.983333</td>
<td>63.25</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error: MS(Error)</td>
<td>30</td>
<td>27.500000</td>
<td>0.916667</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SAS statements for getting variance component estimates:

/* Alternatively, REML estimates of variance components are produced by the MIXED procedure in SAS. Note that there are no terms on the right of the equal sign in the model statement because the only non-random effect is the intercept. */

proc mixed data=set1;
  class batch sample rep;
  model y = ;
  random batch sample(batch);
run;

The Mixed Procedure

Model Information

Data Set WORK.SET1
Dependent Variable y
Covariance Structure Variance Components
Estimation Method REML
Residual Variance Method Profile
Fixed Effects SE Method Model-Based
Degrees of Freedom Method Containment

Class Level Information

Class Levels Values
batch 15 1 2 3 4 5 6 7 8 9 10 11 12 13
      14 15
sample 2 1 2
rep 2 1 2

Dimensions

Covariance Parameters 3
Columns in X 1
Columns in Z 45
Subjects 1
Max Obs Per Subject 60

Covariance Parameter Estimates

Cov Parm        Estimate
batch           7.1280
sample(batch)   28.5333
Residual        0.9167