Nesting and Mixed Effects: Part I
One-way ANOVA
- fixed effects, global test, contrasts, ...

Multiple treatment factors
- factorial treatment structure (fixed effects), two-way ANOVA (or more factors), concept of interaction, 2^k-designs, ...

RCB with factorial treatment structure, ...
- random effects, mixed effects models, nested factor structure, ...

RCB
- random effects, variance components, ...

Block Designs
- inhomogeneous
- one block f., two (more)

Experimental units
- homogeneous

CRD
- one-way ANOVA

One treatment factor

Multiple treatment factors, varied / randomized on different "scales"

Youden Squares

Similar to Lawson (2015)
Where do we stand?

- So far:
  - Fixed effects
  - Random effects
  - Both in the factorial context

- Now:
  - Nested factor structure
  - Mixed models: a combination of fixed and random effects.
With crossed factors $A$ and $B$ we see (by definition) all possible combinations of factor levels, i.e. we can set up a data table of the following form:

<table>
<thead>
<tr>
<th>Factor $A$ / Factor $B$</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>2</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>3</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>4</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
</tbody>
</table>

This means: We see every level of factor $A$ at every level of factor $B$ (and vice versa).

Factor level 1 of factor $A$ has the same meaning across all levels of factor $B$. 
Example: Student Performance (Roth, 2013)

- Want to analyze student performance.

- Data from different classes from different schools (on student level).

- How large is (grade) variability
  - between different schools?
  - between classes within the same school?
  - between students within the same class?

- This looks like a new design, as classes are clearly not crossed with schools, similarly for students.

- This leads us to a new definition…
We call factor $B$ nested in factor $A$ if we have different levels of $B$ within each level of $A$.

- E.g., think of $A = \text{school}$, $B = \text{class}$.
- We also write: $B(A)$.
- Data is not necessarily presented in this form…
# Nested Factors: Example (Roth, 2013)

- **Presented data:**

<table>
<thead>
<tr>
<th>School</th>
<th>Class 1</th>
<th>Class 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>School 1</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>School 2</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>School 3</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

- **Underlying data structure:**

<table>
<thead>
<tr>
<th>School</th>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 3</th>
<th>Class 4</th>
<th>Class 5</th>
<th>Class 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>School 1</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School 2</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>School 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

Hence, **class is nested in school** because class 1 in school 1 has nothing to do with class 1 in school 2 etc.
Nested Factors

- Note: Just because classes are labelled 1 and 2 doesn’t mean that it is a crossed design!

- Hence: Always ask yourself whether the factor level “1” really corresponds to the same “object” across all levels of the other factor.

- Typically we use parentheses in the index to indicate nesting, i.e. the model is written as

\[ Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{k(ij)} \]

- Here, we also wrote the errors in “nested notation”. Errors are always nested, we have just ignored this so far.
Why Use Nesting?

- Typically we use a nested structure due to practical / logistical constraints.

- For example:
  - Patients are nested in hospitals as we don’t want to send patients to all clinics across the country.
  - Samples are nested in batches (in quality control).
  - …
Example: Fully Nested Design

- We call a design **fully nested** if every factor is **nested** in its **predecessor**.

- Genomics example in Oehlert (2010):
  - Consider three subspecies.
  - Randomly choose five males from each subspecies (= 15 males).
  - Each male is mated with four different females of the same subspecies (= 60 females).
  - Observe three offspring per mating (= 180 offsprings).
  - Make two measurements per offspring (= 360 measurements).

- Picture:
Example: Fully Nested Design

- We use the model

\[ Y_{ijklm} = \mu + \alpha_i + \beta_{j(i)} + \gamma_{k(ij)} + \delta_{l(ijk)} + \epsilon_{m(ijkl)} \]

- To calculate the corresponding sums of squares, we use the following decomposition

\[ (y_{ijklm} - \bar{y}_{...}) = (\bar{y}_{i...} - \bar{y}_{...}) + (\bar{y}_{ij...} - \bar{y}_{i...}) + \]

\[ (\bar{y}_{ijk..} - \bar{y}_{ij..}) + (\bar{y}_{ijkl} - \bar{y}_{ijk..}) + (y_{ijklm} - \bar{y}_{ijkl}) \]

take the square and the sum over all indices.
ANOVA Table for Fully Nested Design (Balanced Design)

- This leads us to the decomposition

\[ SS_{Total} = SS_A + SS_{B(A)} + SS_{C(AB)} + SS_{D(ABC)} + SS_E \]

- Assuming we have only random effects and a balanced design we have the following ANOVA table

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>E[MS]</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>( a - 1 )</td>
<td>( \sigma^2 + n\sigma_\delta^2 + nd\sigma_\gamma^2 + ncd\sigma_\beta^2 + nbc\sigma_\alpha^2 )</td>
</tr>
<tr>
<td>B((A))</td>
<td>( a(b - 1) )</td>
<td>( \sigma^2 + n\sigma_\delta^2 + nd\sigma_\gamma^2 + ncd\sigma_\beta^2 )</td>
</tr>
<tr>
<td>C((AB))</td>
<td>( ab(c - 1) )</td>
<td>( \sigma^2 + n\sigma_\delta^2 + nd\sigma_\gamma^2 )</td>
</tr>
<tr>
<td>D((ABC))</td>
<td>( abc(d - 1) )</td>
<td>( \sigma^2 + n\sigma_\delta^2 )</td>
</tr>
<tr>
<td>Error</td>
<td>( abcd(n - 1) )</td>
<td>( \sigma^2 )</td>
</tr>
</tbody>
</table>

- With this information we can again construct tests and estimators for the different variance components.
ANOVA for Fully Nested Designs

- $F$-tests are constructed by taking the ratio of “neighboring” mean squares as they just differ by the variance component of interest.

- This means that we always use the mean square of the successor in the hierarchy tree as denominator.

- E.g., use $F = \frac{MS_A}{MS_{B(A)}}$ to test $H_0: \sigma^2_\alpha = 0$ vs. $H_A: \sigma^2_\alpha > 0$. 

“between $A$”

“within $A$”
Example: Pastes Strength

- Dataset from the \texttt{lme4} package.
- Chemical paste product contained in casks.
- Ten \textbf{deliveries (batches)} were \textit{randomly} selected.
- From each delivery, three \textbf{casks} were \textit{randomly} selected.
- Per cask: make two \textbf{measurements}. 
Example: Pastes Strength - Visualization
Example: Pastes Strength

- **Model:**

\[ Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{k(ij)} \]

- **Dataset in R:**

```r
> str(Pastes)
'data.frame': 60 obs. of 4 variables:
$ strength: num 62.8 62.6 60.1 62.3 62.7 63.1 60 61.4 57.5
$ batch : Factor w/ 10 levels "A","B","C","D",..: 1 1 1 1 1 1 1 1 1
$ cask  : Factor w/ 3 levels "a","b","c": 1 1 2 3 3 1 1
$ sample: Factor w/ 30 levels "A:a","A:b","A:c",..: 1 1 2
```

- **Be careful! Why?**
Analysis Using lmer

> fm1 <- lmer(strength ~ (1 | batch/cask), data = Pastes)
> summary(fm1)

Linear mixed model fit by REML

Formula: strength ~ (1 | batch/cask)

Data: Pastes

REML criterion at convergence: 247

Scaled residuals:

    Min  1Q Median  3Q Max
-1.4798 -0.5156  0.0095  0.4720  1.3897

Random effects:

<table>
<thead>
<tr>
<th>Groups</th>
<th>Name</th>
<th>Variance</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>cask:batch</td>
<td>(Intercept)</td>
<td>8.434</td>
<td>2.9041</td>
</tr>
<tr>
<td>batch</td>
<td>(Intercept)</td>
<td>1.657</td>
<td>1.2874</td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>0.678</td>
<td>0.8234</td>
</tr>
</tbody>
</table>

Number of obs: 60, groups: cask:batch, 30; batch, 10

Fixed effects:

|                  | Estimate | Std. Error | df | t value | Pr(>|t|) |
|------------------|----------|------------|----|---------|----------|
| (Intercept)      | 60.0533  | 0.6769     | 9.0000 | 88.72   | 1.49e-14 ***|

See alternative formulations in the corresponding R-file.
Analysis Using lmer

- Get **confidence intervals** using
  ```r
  > confint(fm1, oldNames = FALSE)
  Computing profile confidence intervals ...
   2.5%   97.5%
  sd_(Intercept)|cask:batch  2.1579337  4.053589
  sd_(Intercept)|batch        0.0000000  2.946591
  sigma          0.6520234  1.085448
  (Intercept)     58.6635504  61.443016
  ```

- (Conservative) tests of the variance components can here be obtained with
  ```r
  > fm2 <- lmer(strength ~ (1 | batch) + (1 | cask:batch), data = Pastes)
  > rand(fm2)
  Analysis of Random effects Table:
   Chi.sq Chi.DF p.value
  batch    0.658    1   0.4
  cask:batch 54.605    1 1e-13 ***
  ```

- **Alternative:** Package RLRsim or bootstrap based methods (will not discuss this here).
Analysis Using `aov`

- **Using the `aov` command**
  ```r
  fit <- aov(strength ~ batch + cask %in% batch, data = Pastes)
  summary(fit)
  ```
  ```
  Df Sum Sq Mean Sq F value Pr(>F)
  batch  9 247.4  27.489  40.55 2.28e-14  ***
  batch:cask 20 350.9  17.545  25.88 9.79e-14  ***
  Residuals 30  20.3   0.678
  ```

- Output can be used to manually calculate the different variance components by solving the equations with the corresponding expected mean squares:

  \[
  \hat{\sigma}^2 = 0.678 \\
  \hat{\sigma}^2_\beta = \frac{17.545 - 0.678}{2} = 8.43 \\
  \hat{\sigma}^2_\alpha = \frac{27.489 - 17.545}{2 \cdot 3} = 1.66
  \]

<table>
<thead>
<tr>
<th>E[MS]</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>batch</td>
<td>(\sigma^2 + 2 \cdot \sigma^2_\beta + 2 \cdot 3 \cdot \sigma^2_\alpha)</td>
</tr>
<tr>
<td>cask(batch)</td>
<td>(\sigma^2 + 2 \cdot \sigma^2_\beta)</td>
</tr>
<tr>
<td>Error</td>
<td>(\sigma^2)</td>
</tr>
</tbody>
</table>
Analysis Using aov

- Similarly, tests have to be calculated **manually**.

- E.g., for the batch variance component:

  \[ F = \frac{27.489}{17.545} = 1.567 \]

- Use \( F_{9,20}\)-distribution to calculate \( p \)-value:

  ```r
  > pf(27.489 / 17.545, df1 = 9, df2 = 20, lower.tail = FALSE)
  [1] 0.1925487
  ```

- Hence, **cannot** reject \( H_0: \sigma_\alpha^2 = 0 \).

- Note that default output is **wrong** here as the model was interpreted as a fixed effects model (using the wrong denominator mean square)!
General Situation

- The fully nested design is only a (very) special case.
- A design can of course have both crossed and nested factors.
- In addition, a model can contain both random and fixed effects.
- If this is the case, we call it a mixed effects model.
- Let’s have a look at an example.
Cheese Tasting (Oehlert, 2010, Example 12.2)

- How do urban and rural consumers rate cheddar cheese for bitterness?

- Four 50-pound blocks of different cheese types are available.

- We use food science students as our raters
  - Choose ten students at random with rural background.
  - Choose ten students at random with urban background.

- Each rater will taste eight bites of cheese (presented in random order).

- The eight bites consist of two from each cheese type. Hence, every rater gets every cheese type twice.
What factors do we have here?
- \( A: \text{background} \), levels = \{“rural”, “urban”\}
- \( B: \text{rater} \), levels = \{1, \ldots, 10\} (or 20); \text{nested} in background
- \( C: \text{cheese type} \), levels = \{1, 2, 3, 4\}

Relationship: \( A \times B (A) \times C \) (both \( A \) and \( B \) are crossed with \( C \))

Cheese Tasting (Oehlert, 2010, Example 12.2)
Cheese Tasting (Oehlert, 2010, Example 12.2)

- A model to analyze this data could be

\[ Y_{ijkl} = \mu + \alpha_i + \beta_{j(i)} + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk(i)} + \epsilon_{l(ijk)} \]

where

- \(\alpha_i\) are the fixed effects of background
- \(\beta_{j(i)}\) are the random effects of rater (within background)
- \(\gamma_k\) are the fixed effects of cheese type.
- \((\alpha\gamma)_{ik}\) is the (fixed) interaction effect between background and cheese type.
- \((\beta\gamma)_{jk(i)}\) is the (random) interaction between rater and cheese type.

- The interaction between a fixed effect and a random effect is random (as it includes a random component).
Cheese Tasting (Oehlert, 2010, Example 12.2)

- Interpretation of parameters:

<table>
<thead>
<tr>
<th>Term</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_i$</td>
<td>Main effect of background.</td>
</tr>
<tr>
<td>$\beta_{j(i)}$</td>
<td>Random effect of rater: allows for an individual “general cheese liking” level of a rater.</td>
</tr>
<tr>
<td>$\gamma_k$</td>
<td>Main effect of cheese type.</td>
</tr>
<tr>
<td>$(\alpha\gamma)_{ik}$</td>
<td>Fixed interaction effect between background and cheese type: allows for a background specific cheese type preference.</td>
</tr>
<tr>
<td>$(\beta\gamma)_{jk(i)}$</td>
<td>Random interaction between rater and cheese type: allows for an individual deviation from the population average “cheese type” effect.</td>
</tr>
</tbody>
</table>

- See R-File for results (later).
Appendix: Data Generating Mechanisms

- Assume only two factors:
  - \( A \) fixed (with \( a \) levels) and
  - \( B \) random (with \( b \) levels)

- Think of hypothetical data-table

<table>
<thead>
<tr>
<th>A/B</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( a )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

with \textit{lot's} of columns.

- To get the observed data-table we randomly pick out \( b \) columns.
Appendix: Data Generating Mechanisms

- That means: if we repeat the experiment and select the same column twice we get the very same column (and of course the same column total).

- This implicitly means: the interaction effects are “attached” to the column. This is called the **restricted model**.

- In the restricted model we assume that the interaction effects add to zero when summed across a fixed effect (they are random but **restricted**!)

- The alternative is the **unrestricted model** which treats interaction effects independently from the main effects.

- `lmer` uses the unrestricted model.
Appendix: Data Generating Mechanisms

- In the tasting experiment: Would prefer restricted model because interaction is “attached” to raters.

- Reason: the rater will not change his special taste (remember meaning of parameters).
Appendix: Data Generating Mechanisms: More Technical

- **Unrestricted model:**
  - Random effects (including interactions!) have the assumptions:
    - independent
    - normally distributed with mean 0
    - effects corresponding to the same term have a common variance: $\sigma^2_\alpha, \sigma^2_\beta, \sigma^2_{\alpha\beta}$ etc.

  **Fixed effects:**
  - have the usual sum-to-zero constraint (across any subscript).

- **Restricted model:**
  - As above, with the exception that interactions between random and fixed factors (which are random!) follow the sum-to-zero constraint over any subscript corresponding to a fixed factor.
  - This induces a negative correlation within these random effects, hence they are not independent anymore.