Random Effects

Lukas Meier, Seminar für Statistik
New Philosophy…

- Up to now: treatment effects were **fixed, unknown parameters** that we were trying to **estimate**.
- Such models are also called **fixed effects models**.
- Now: Consider the situation where treatments are **random samples** from a **large population** of potential treatments.
- Example: Effect of machine operators that were randomly selected from a large pool of operators.
- In this setup, treatment effects are **random variables** and therefore called **random effects**. The corresponding model will be a **random effects model**.
Why would we be interested in a random effects situation?

It is a useful way of thinking if we want to make a statement (conclusion) about the population of all treatments.

In the operator example we shift the focus away from the individual operators (treatments) to the population of all operators (treatments).

Typically, we are interested in the variance of the treatment population.

E.g., what is the variation from operator to operator?
Examples of Random Effects

<table>
<thead>
<tr>
<th>Randomly select...</th>
<th>...from...</th>
</tr>
</thead>
<tbody>
<tr>
<td>clinics</td>
<td>...all clinics in a country.</td>
</tr>
<tr>
<td>school classes</td>
<td>...all school classes in a region.</td>
</tr>
<tr>
<td>investigators</td>
<td>...a large pool of investigators.</td>
</tr>
<tr>
<td>series in quality control</td>
<td>...all series in a certain time period.</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Carton Experiment One (Oehlert, 2000)

- Company with 50 machines that produce cardboard cartons.
- Ideally, strength of the cartons shouldn’t vary too much.
- Therefore, we want to have an idea about
  - “machine-to-machine” variation
  - “sample-to-sample” variation on the same machine.
- Perform experiment:
  - Choose 10 machines at random (out of the 50)
  - Produce 40 cartons on each machine
  - Test resulting cartons for strength (→ response)
Carton Experiment One (Oehlert, 2000)

- Model so far:
  \[ Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \]
  where \( \alpha_i \) is the (fixed) effect of machine \( i \) and \( \epsilon_{ij} \) are the errors with the usual assumptions.

- However, this model does not reflect the sampling mechanism from above.

- If we repeat the experiment, the selected machines change and therefore also the meaning of the parameters: they typically correspond to a different machine!

- Moreover, we want to learn something about the population of all machines.
Carton Experiment One (Oehlert, 2000)

- **New: Random effects model:**

  \[ Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \]

  with
  - \( \alpha_i \) i.i.d. \( \sim N(0, \sigma^2_\alpha) \) (NEW)
  - \( \epsilon_{ij} \) i.i.d. \( \sim N(0, \sigma^2) \)

- This looks very similar to the old model, however the \( \alpha_i \)'s are now **random variables**!

- That small change will have a **large impact** on the properties of the model and on our way to analyze such kind of data.
Carton Experiment One (Oehlert, 2000)

- Properties of random effects model:
  - $\text{Var}(Y_{ij}) = \sigma^2_\alpha + \sigma^2$
  - $\text{Cor}(Y_{ij}, Y_{kl}) = \begin{cases} 
  0 & i \neq k \\
  \sigma^2_\alpha / (\sigma^2_\alpha + \sigma^2) & i = k, j \neq l \\
  1 & i = k, j = l 
\end{cases}$

Reason: Observations from the same machine “share” the same random value $\alpha_i$ and are therefore correlated.

- Conceptually, we could also put all the correlation structure into the error term and forget about the $\alpha_i$’s, i.e.

$$Y_{ij} = \mu + \epsilon_{ij}$$

where $\epsilon_{ij}$ has the appropriate correlation structure from above. Sometimes this interpretation is a useful way of thinking.
Random vs. Fixed: Overview

- **Comparison** between random and fixed effects models

<table>
<thead>
<tr>
<th>Term</th>
<th>Fixed effects model</th>
<th>Random effects model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_i$</td>
<td>fixed, unknown constant</td>
<td>$\alpha_i$ i.i.d. $\sim N(0, \sigma^2_\alpha)$</td>
</tr>
<tr>
<td>Side constraint on $\alpha_i$</td>
<td>needed</td>
<td>not needed</td>
</tr>
<tr>
<td>$E[Y_{ij}]$</td>
<td>$\mu + \alpha_i$</td>
<td>$\mu$, but $E[Y_{ij}</td>
</tr>
<tr>
<td>$\text{Var}(Y_{ij})$</td>
<td>$\sigma^2$</td>
<td>$\sigma^2_\alpha + \sigma^2$</td>
</tr>
<tr>
<td>$\text{Corr}(Y_{ij}, Y_{kl})$</td>
<td>$= 0 (j \neq l)$</td>
<td>$= \left{ \begin{array}{ll} 0 &amp; i \neq k \ \sigma^2_\alpha/(\sigma^2_\alpha + \sigma^2) &amp; i = k, j \neq l \ 1 &amp; i = k, j = l \end{array} \right.$</td>
</tr>
</tbody>
</table>

- **A note on the sampling mechanism:**
  - Fixed: Draw new random errors only, everything else is kept **constant**.
  - Random: Draw new “treatment effects” **and** new random errors (!)
Illustration of Correlation Structure

Fixed case: 3 different fixed treatment levels $\alpha_i$.
We (repeatedly) sample 2 observations per treatment level:

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

Think of 3 specific machines

Think of 2 carton samples

$\alpha_1 = -4.5$
$\alpha_2 = 1$
$\alpha_3 = 3.5$
Illustration of Correlation Structure

Random case:
Whenever we draw 2 observations $Y_{i1}$ and $Y_{i2}$ we first have to draw a **new** (common) random treatment effect $\alpha_i$. 

Think of a random machine.
Carton Experiment Two (Oehlert, 2000)

- Let us extend the previous experiment.
- Assume that machine operators also influence the production process.
- Choose 10 operators at random.
- Each operator will produce 4 cartons on each machine (hence, operator and machine are crossed factors).
- All assignments are completely randomized.
Carton Experiment Two (Oehlert, 2000)

- Model:

\[ Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}, \]

with

- \( \alpha_i, \beta_j, (\alpha\beta)_{ij}, \epsilon_{ijk} \) independent and normally distributed.
- \( \text{Var}(Y_{ijk}) = \sigma^2_\alpha + \sigma^2_\beta + \sigma^2_{\alpha\beta} + \sigma^2 \) (different variance components).
- Measurements from the same machine and / or operator are again correlated.
- The more random effects two observations share, the larger the correlation. It is given by

\[
\frac{\text{sum of shared variance components}}{\text{sum of all variance components}}
\]

- E.g., correlation between two (different) observations from the same operator on different machines is given by

\[
\frac{\sigma^2_\beta}{\sigma^2_\alpha + \sigma^2_\beta + \sigma^2_{\alpha\beta} + \sigma^2}
\]
Carton Experiment Two (Oehlert, 2000)

- **Hierarchy** is typically less problematic in random effects models.
  1) What part of the variation is due to general machine-to-machine variation? → $\sigma^2_\alpha$
  2) What part of the variation is due to operator-specific machine variation? → $\sigma^2_{\alpha\beta}$

Could ask question (1) even if interaction is present (question (2)).

- Extensions to **more than two factors** straightforward.
ANOVA for Random Effects Models (balanced designs)

- Sums of squares, degrees of freedom and mean squares are being calculated as if the model would be a **fixed effects** model (!)

- One-way ANOVA ($A$ random, $n$ observations per cell)

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>$E[MS]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>$g - 1$</td>
<td>...</td>
<td>...</td>
<td>$\sigma^2 + n\sigma^2_{\alpha}$</td>
</tr>
<tr>
<td>Error</td>
<td>$N - g$</td>
<td>...</td>
<td>...</td>
<td>$\sigma^2$</td>
</tr>
</tbody>
</table>

- Two-way ANOVA ($A$, $B$, $AB$ random, $n$ observations per cell)

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>$E[MS]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>$a - 1$</td>
<td>...</td>
<td>...</td>
<td>$\sigma^2 + b \cdot n \cdot \sigma^2_{\alpha} + n \cdot \sigma^2_{\alpha\beta}$</td>
</tr>
<tr>
<td>$B$</td>
<td>$b - 1$</td>
<td>...</td>
<td>...</td>
<td>$\sigma^2 + a \cdot n \cdot \sigma^2_{\beta} + n \cdot \sigma^2_{\alpha\beta}$</td>
</tr>
<tr>
<td>$AB$</td>
<td>$(a - 1)(b - 1)$</td>
<td>...</td>
<td>...</td>
<td>$\sigma^2 + n \cdot \sigma^2_{\alpha\beta}$</td>
</tr>
<tr>
<td>Error</td>
<td>$ab(n - 1)$</td>
<td>...</td>
<td>...</td>
<td>$\sigma^2$</td>
</tr>
</tbody>
</table>
One-Way ANOVA with Random Effects

- We are now formulating our null-hypothesis with respect to the parameter $\sigma^2_{\alpha}$.

- To test $H_0: \sigma^2_{\alpha} = 0$ vs. $H_A: \sigma^2_{\alpha} > 0$ we use the ratio
  \[ F = \frac{MS_A}{MS_E} \sim F_{g-1,N-g} \] under $H_0$

  **Exactly** as in the fixed effect case!

- Why? Under the **old and the new** $H_0$ both models are the same!
Two-Way ANOVA with Random Effects

- To test $H_0: \sigma^2_{\alpha} = 0$ we need to find a term which has identical $E[MS]$ under $H_0$.

- Use $MS_{AB}$, i.e. $F = \frac{MS_A}{MS_{AB}} \sim F_{a-1, (a-1)(b-1)}$ under $H_0$.

- Similarly for the test $H_0: \sigma^2_{\beta} = 0$.

- The interaction will be tested against the error, i.e. use

  $$F = \frac{MS_{AB}}{MS_E} \sim F_{(a-1)(b-1), ab(n-1)}$$

  under $H_0: \sigma^2_{\alpha\beta} = 0$.

- In the fixed effect case we would test all effects against the error term (i.e., use $MS_E$ instead of $MS_{AB}$ to build $F$-ratio)!
Two-Way ANOVA with Random Effects

- Reason: ANOVA table for fixed effects:

<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>$\sigma^2 + b \cdot n \cdot Q(\alpha)$</td>
</tr>
<tr>
<td>$B$</td>
<td>$b - 1$</td>
<td>$\sigma^2 + a \cdot n \cdot Q(\beta)$</td>
</tr>
<tr>
<td>$AB$</td>
<td>$(a - 1)(b - 1)$</td>
<td>$\sigma^2 + n \cdot Q(\alpha\beta)$</td>
</tr>
<tr>
<td>Error</td>
<td>$ab(n - 1)$</td>
<td>$\sigma^2$</td>
</tr>
</tbody>
</table>

- E.g, $SS_A (MS_A)$ is being calculated based on column-wise means.

- In the **fixed effects model**, the expected mean squares do **not** “contain” any other component.
Two-Way ANOVA with Random Effects

- In a random effects model, a column-wise mean is “contaminated” with the average of the corresponding interaction terms.

- In a fixed effects model, the sum (or mean) of these interaction terms is zero by definition.

- In the random effects model, this is only true for the expected value, but not for an individual realization!

- Hence, we need to check whether the variation from “column to column” is larger than term based on error and interaction term.
We do not only want to test the variance components, we also want to have estimates of them. I.e., we want to determine $\hat{\sigma}_\alpha^2, \hat{\sigma}_\beta^2, \hat{\sigma}_{\alpha\beta}^2, \hat{\sigma}^2$ etc.

Easiest approach: ANOVA estimates of variance components.

Use columns “MS” and “E[MS]” in ANOVA table, solve the corresponding equations from bottom to top.

Example: One-way ANOVA

- $\hat{\sigma}^2 = MS_E$
- $\hat{\sigma}_\alpha^2 = \frac{(MS_A - MS_E)}{n}$
Point Estimates of Variance Components

- Advantage: Can be done using **standard ANOVA functions** (i.e., no special software needed).

- Disadvantages:
  - Estimates can be negative (in previous example if $MS_A < MS_E$). Set them to zero in such cases.
  - Not always as easy as here.

- This is like a method of moments estimator.

- More modern and much more flexible: **restricted maximum-likelihood estimator** (REML).
Point Estimates of Variance Components: REML

- Think of a modification of maximum likelihood estimation that removes bias in estimation of variance components.
- Theory complicated (still ongoing research).
- Software implementation in R-package lme4 (or lmerTest)

lme4 and lmerTest allow to fit so called mixed models (containing both random and fixed effects, more details later).

Basically, lmerTest is the same as lme4 with some more features.
Confidence Intervals for Variance Components

- General rule: Variances are “difficult” to estimate in the sense that you’ll need a lot of observations to have some reasonable accuracy.
- Only approximate confidence intervals are available.
- Use confint in R.
Some Thoughts About Random Effects

- If we do a study with random effects it is good if we have a lot of levels of a random effect in order to estimate a variance component with high precision.

- Or in other words: Who wants to estimate a variance with only very few observations?
Example: Genetics Study (Kuehl, 2000, Exercise 5.1)

- Genetics study with **beef animals**.
- **Inheritance study of birth weights**.
- Five **sires**, each mated to a different group of **dams**.
- Birth weight of eight male calves in each of the five sire groups.

<table>
<thead>
<tr>
<th>Sire</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>61</td>
<td>100</td>
<td>56</td>
<td>113</td>
<td>99</td>
<td>103</td>
<td>75</td>
<td>62</td>
</tr>
<tr>
<td>2</td>
<td>75</td>
<td>102</td>
<td>95</td>
<td>103</td>
<td>98</td>
<td>115</td>
<td>98</td>
<td>94</td>
</tr>
<tr>
<td>3</td>
<td>58</td>
<td>60</td>
<td>60</td>
<td>57</td>
<td>57</td>
<td>59</td>
<td>54</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>57</td>
<td>56</td>
<td>67</td>
<td>59</td>
<td>58</td>
<td>12</td>
<td>101</td>
<td>101</td>
</tr>
<tr>
<td>5</td>
<td>59</td>
<td>46</td>
<td>120</td>
<td>115</td>
<td>115</td>
<td>93</td>
<td>105</td>
<td>75</td>
</tr>
</tbody>
</table>

- Analyze data using a **random effect for sire**.
Example: Genetics Study  *(Kuehl, 2000, Chapter 5, Ex. 1)*
Example: Genetics Study

- Model: \( Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \alpha_i \text{ i.i.d. } \sim N(0, \sigma^2_{\alpha}), \epsilon_{ij} \text{ i.i.d. } \sim N(0, \sigma^2) \)

\[
> \text{fit} <- \text{aov(weight ~ sire, data = animals)} \\
> \text{summary(fit)}
\]

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>sire</td>
<td>4</td>
<td>5591</td>
<td>1397.8</td>
<td>3.014</td>
<td>0.0309</td>
</tr>
<tr>
<td>Residuals</td>
<td>35</td>
<td>16233</td>
<td>463.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- We reject \( H_0: \sigma^2_{\alpha} = 0 \).

- We estimate \( \sigma^2_{\alpha} \) by \( \hat{\sigma}^2_{\alpha} = \frac{1397.8 - 463.8}{8} = 116.75 \).

- The variance of \( Y_{ij} \) is estimated as
  \[
  \hat{\sigma}^2 + \hat{\sigma}^2_{\alpha} = 116.75 + 463.8 = 580.55.
  \]

- Variation due to sire accounts for about 20% of total variance (= intraclass correlation).

Old school estimation technique.
Example: Genetics Study

- We fitted the model as if it was a fixed effects model and then “adjusted” the output for random effects specific questions.

- Now we want to use the more modern approach (based on REML estimation technique).
Example: Genetics Study

- In R using the function `lmer` in Package `lme4`.

```r
> fit.lme <- lmer(weight ~ 1 | sire, data = animals)
> summary(fit.lme)
Linear mixed model fit by REML ['lmerMod']
Formula: weight ~ 1 | sire
    Data: animals

REML criterion at convergence: 358.2

Scaled residuals:
          Min       1Q   Median       3Q      Max
-1.9593  -0.7459  -0.1581   0.8143   1.9421

Random effects:
 Groups     Name   Variance  Std.Dev.
    sire   (Intercept) 116.7     10.81
    Residual          463.8     21.54
Number of obs: 40, groups: sire, 5

Fixed effects:
                     Estimate  Std. Error t value
(Intercept)       82.5500     5.9111  13.96
```

- Meaning: a random effect per sire
- \( \hat{\sigma}_\alpha \)
- \( \hat{\sigma} \)
- Check if model was interpreted correctly
- \( \hat{\mu} \)
Example: Evaluating Machine Performance  
(Kuehl, 2000, Ex. 7.1)

- Manufacturer was developing a new spectrophotometer for medical labs.

- Development at pilot stage. Evaluate machine performance from assembly line production.

- Critical: **Consistency** of measurement from **day to day** among **different machines**.

- Design:
  - 4 (randomly selected) machines
  - 4 (randomly selected) days

- Per day: 8 serum samples (from the **same** stock reagent), randomly assign 2 samples to each of the 4 machines.
Example: Evaluating Machine Performance

- Measure triglyceride levels (mg/dl) of the samples.

- Note: Always the same technician prepared the serum samples and operated the machines throughout the experiment.
Example: Evaluating Machine Performance

- Fit random effects model with interaction with usual assumpt.

\[ Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \]

- Classical approach:
  ```r
  fit <- aov(y ~ day * machine, data = trigly)
  summary(fit)
  ```

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>day</td>
<td>3</td>
<td>1334.5</td>
<td>444.8</td>
<td>24.86</td>
<td>2.91e-06 ***</td>
</tr>
<tr>
<td>machine</td>
<td>3</td>
<td>1647.3</td>
<td>549.1</td>
<td>30.68</td>
<td>7.19e-07 ***</td>
</tr>
<tr>
<td>day:machine</td>
<td>9</td>
<td>786.0</td>
<td>87.3</td>
<td>4.88</td>
<td>0.00294 **</td>
</tr>
<tr>
<td>Residuals</td>
<td>16</td>
<td>286.3</td>
<td>17.9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- “Classical” approach to estimate variance components.

- Results:

\[ \hat{\sigma}^2 = 17.9 \]

\[ \hat{\sigma}^2 = \frac{87.3 - 17.9}{2} = 34.7 \]

\[ \hat{\sigma}_\alpha^2 = \frac{444.8 - 87.3}{8} = 44.7 \]

\[ \hat{\sigma}_\beta^2 = \frac{549.1 - 87.3}{8} = 57.7 \]
Example: Evaluating Machine Performance

Testing the variance components: “by hand”

- **Interaction:** $H_0: \sigma_{\alpha\beta}^2 = 0.$

  \[
  \frac{MS_{AB}}{MS_E} = \frac{87.3}{17.9} = 4.9, \ F_{9,16}\text{-distribution}
  \]

  > pf(87.3 / 17.9, 9, 16, lower.tail = FALSE)

  [1] 0.002946051

  → **reject**

- **Main effect day:** $H_0: \sigma_{\alpha}^2 = 0.$

  \[
  \frac{MS_A}{MS_{AB}} = \frac{444.8}{87.3} = 5.1, \ F_{3,9}\text{-distribution}
  \]

  > pf(444.8 / 87.3, 3, 9, lower.tail = FALSE)

  [1] 0.02477665

  → **reject**

- **Main effect machine:** $H_0: \sigma_{\beta}^2 = 0.$

  \[
  \frac{MS_B}{MS_{AB}} = \frac{549.1}{87.3} = 6.3, \ F_{3,9}\text{-distribution}
  \]

  > pf(549.1 / 87.3, 3, 9, lower.tail = FALSE)

  [1] 0.01370686

  → **reject**
Example: Evaluating Machine Performance

- Using the function `lmer` in package `lme4`

```r
> fit.lme <- lmer(y ~ (1 | day) + (1 | machine) + (1 | machine:day), data = trigly)
> summary(fit.lme)
```

Linear mixed model fit by REML ['lmerMod']
Formula: y ~ (1 | day) + (1 | machine) + (1 | machine:day)
Data: trigly

REML criterion at convergence: 215

Scaled residuals:
Min 1Q Median 3Q Max
-1.84282 -0.35581 0.03484 0.20699 2.31766

Random effects:
<table>
<thead>
<tr>
<th>Groups</th>
<th>Name</th>
<th>Variance</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>machine:day</td>
<td>(Intercept)</td>
<td>34.72</td>
<td>5.892</td>
</tr>
<tr>
<td>machine</td>
<td>(Intercept)</td>
<td>57.72</td>
<td>7.597</td>
</tr>
<tr>
<td>day</td>
<td>(Intercept)</td>
<td>44.69</td>
<td>6.685</td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>17.90</td>
<td>4.230</td>
</tr>
</tbody>
</table>

Number of obs: 32, groups: machine:day, 16; machine, 4; day, 4

Fixed effects:
<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>141.184</td>
<td>5.323</td>
<td>26.52</td>
</tr>
</tbody>
</table>

Meaning: a random effect per day, per machine and per day x machine combination.

Check if model was interpreted correctly.
Example: Evaluating Machine Performance

- Total variance is $17.9 + 34.7 + 44.7 + 57.7 = 155$.

- Individual contributions

<table>
<thead>
<tr>
<th>Source</th>
<th>Percentage</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day</td>
<td>$\frac{44.7}{155} = 29%$</td>
<td>Day to day operational differences (e.g., due to daily calibration)</td>
</tr>
<tr>
<td>Machine</td>
<td>$\frac{57.7}{155} = 37%$</td>
<td>Variability in machine performance</td>
</tr>
<tr>
<td>Interaction</td>
<td>$\frac{34.7}{155} = 22%$</td>
<td>Variability due to inconsistent behavior of machines over days (calibration inconsistency within the same day?)</td>
</tr>
<tr>
<td>Error</td>
<td>$\frac{17.9}{155} = 12%$</td>
<td>Variation in serum samples</td>
</tr>
</tbody>
</table>

- Manufacturer now has to decide if some sources of variation are too large.