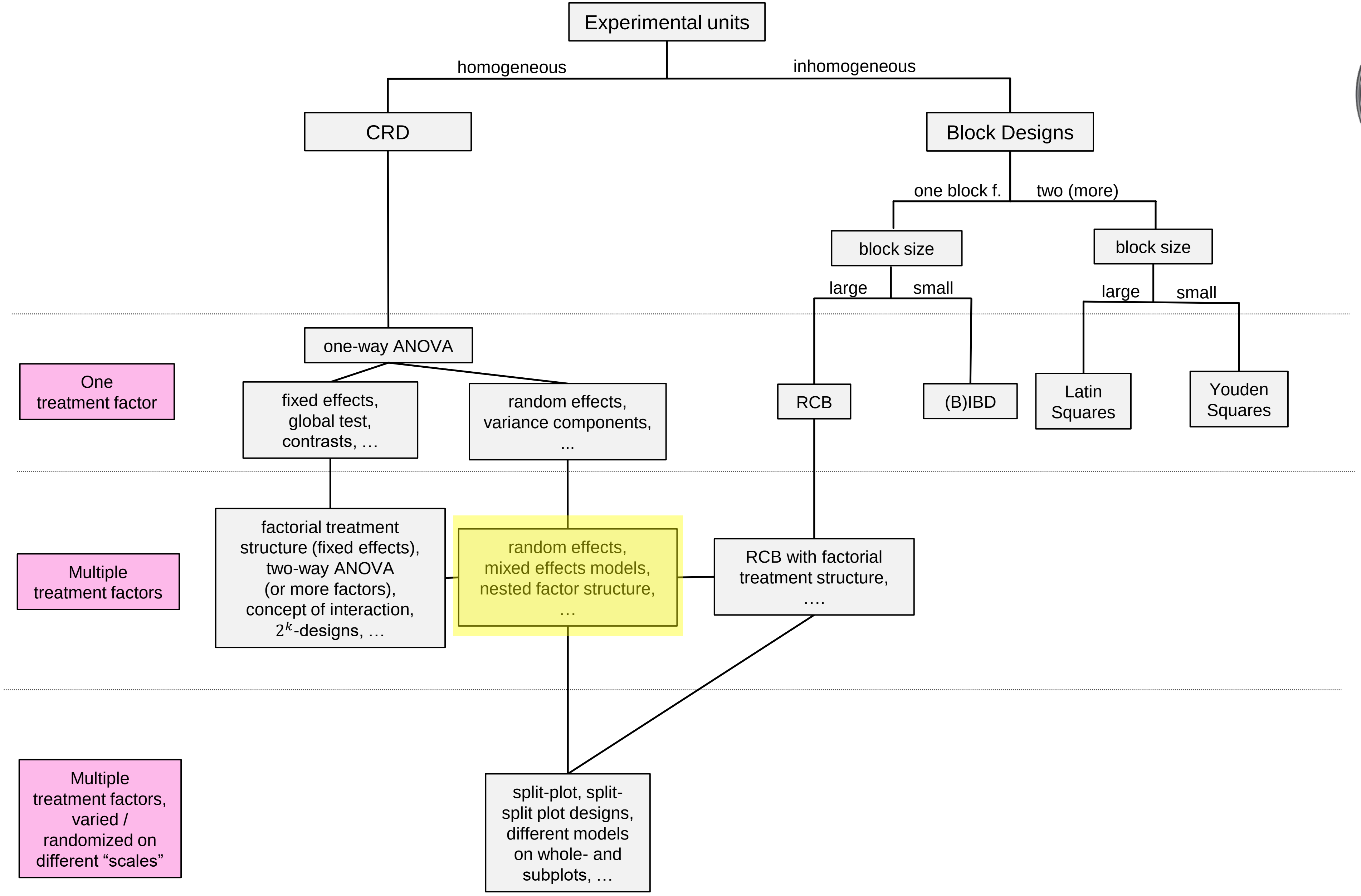
Nesting and Mixed Effects: Part II



Similar to Lawson (2015)

Mixed Effects Models

- Before we do the full analysis for the cheese rating example, we have a look at two easier examples.

- We use them to get a better idea about
  - how to fit such models in R,
  - how to interpret the corresponding parameters,
  - … especially the difference to purely fixed effects models.

Example: Stools

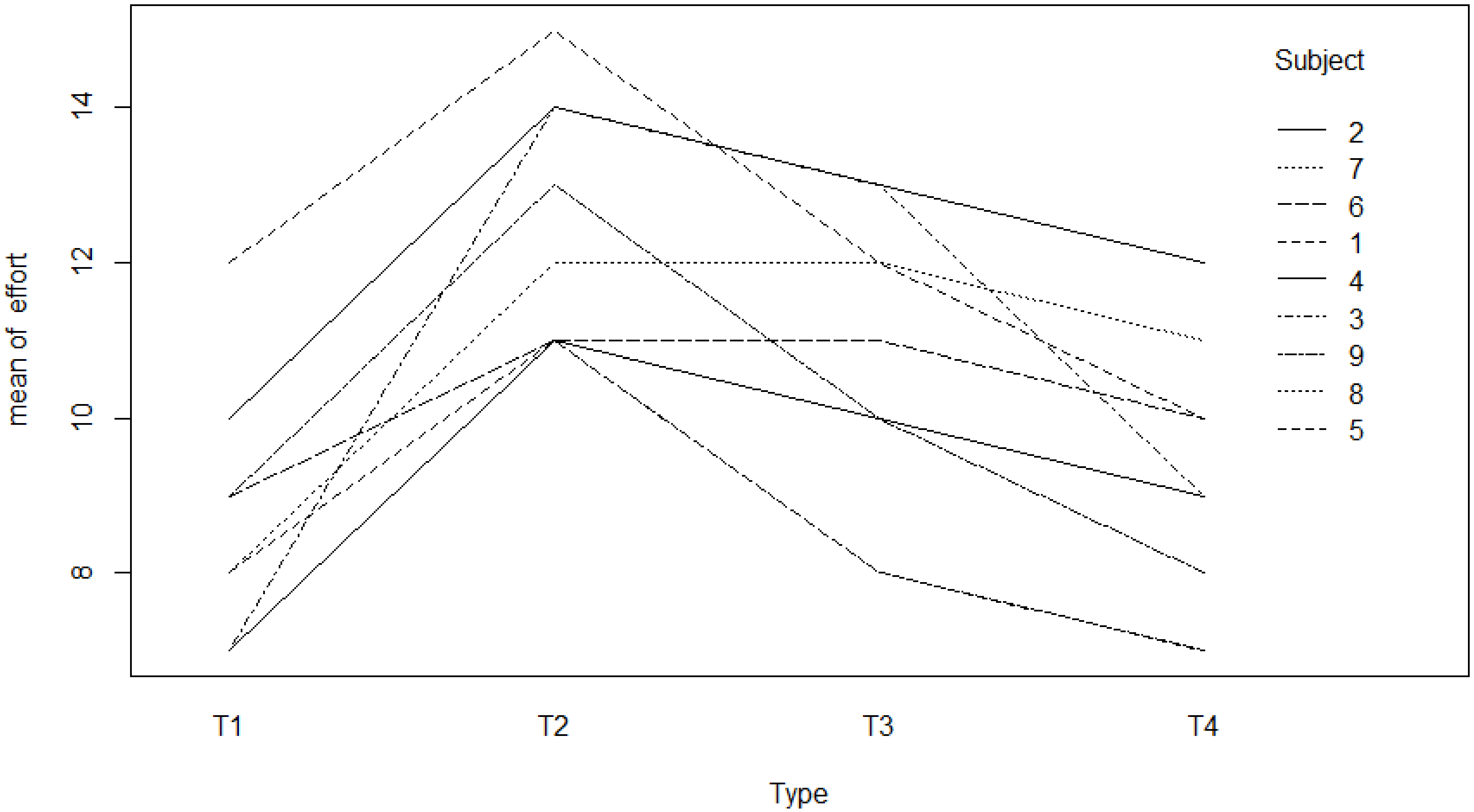
- Dataset `ergoStool` from R-package `nlme`.

- As stated in the help file:

  From an article in Ergometrics (1993, pp. 519-535) on “The Effects of a Pneumatic Stool and a One-Legged Stool on Lower Limb Joint Load and Muscular Activity.”

- Overview of data
  - 4 different **stool types**
  - 9 different **subjects** (randomly selected)
  - 1 measurement per combination of stool type and subject: **Effort** on the so called Borg scale

Example: Stools - Visualization



Example: Stools - Model

- We analyze the data with the following **mixed effects** model:

$$\underbrace{Y_{ij}}_{\text{effort}} = \mu + \underbrace{\alpha_i}_{\text{fixed effect of }\textbf{stool type}} + \overbrace{\underbrace{\beta_j}_{\text{random effect of }\textbf{subject}}}^{N(0,\sigma_\beta^2)} + \overbrace{\underbrace{\epsilon_{ij}}_{\text{error term}}}^{N(0,\sigma^2)}$$

- For the α_i's we have to use a side-constraint (e.g, sum-to-zero or set reference treatment to zero).
- Here, subject is a (random) block factor.
- In R we fit this using the `lmer` function

```
> fit <- lmer(effort ~ Type + (1 | Subject), data = ergoStool)
```

fixed effect of **stool type**

random effect per **subject**

Example: Stools - Output

- The standard summary output looks as follows:

```
> summary(fit)
Linear mixed model fit by REML t-tests use Satterthwaite approximations to degrees of
  freedom [merModLmerTest]
Formula: effort ~ Type + (1 | Subject)
   Data: ergoStool

REML criterion at convergence: 121.1

Scaled residuals:
     Min       1Q   Median       3Q      Max
-1.80200 -0.64317  0.05783  0.70100  1.63142

Random effects:
 Groups   Name        Variance Std.Dev.
 Subject  (Intercept) 1.775    1.332
 Residual             1.211    1.100
Number of obs: 36, groups:  Subject, 9

Fixed effects:
            Estimate Std. Error      df t value Pr(>|t|)
(Intercept)   8.5556     0.5760 15.5300  14.853 1.36e-10 ***
TypeT2        3.8889     0.5187 24.0000   7.498 9.75e-08 ***
TypeT3        2.2222     0.5187 24.0000   4.284 0.000256 ***
TypeT4        0.6667     0.5187 24.0000   1.285 0.210951
```

$\hat{\sigma}_\beta$

$\hat{\sigma}$

$\hat{\mu}$

$\hat{\alpha}_2$

$\hat{\alpha}_3$

$\hat{\alpha}_4$

Coefficients in terms of the “coded” variables. Need to know encoding scheme for interpretation.

Example: Stools - Output

- We can perform the **global** F-test for **stool type** by calling `anova` on the fitted object.

```
> anova(fit)
Analysis of Variance Table of type III  with  Satterthwaite
approximation for degrees of freedom
     Sum Sq Mean Sq NumDF DenDF F.value    Pr(>F)
Type 81.194  27.065     3    24  22.356 3.935e-07 ***
```

- We can also test the **variance component** of **subjects** and calculate the **confidence intervals for all effects** using

```
> rand(fit)
Analysis of Random effects Table:
        Chi.sq Chi.DF p.value
Subject   13.5      1   2e-04 ***
```

conservative test

```
> confint(fit, oldNames = FALSE)
Computing profile confidence intervals ...
                          2.5 %   97.5 %
sd_(Intercept)|Subject  0.7342354 2.287261
sigma                   0.8119798 1.390104
(Intercept)             7.4238425 9.687269
TypeT2                  2.8953043 4.882473
TypeT3                  1.2286377 3.215807
TypeT4                 -0.3269179 1.660251
```

Example: Stools - Interpretation

- Interpretation of previous outputs:
  - **Stool type** is **highly significant** (p-value from global F-test).
  - Stool type effects can be read off from the fixed effects part of the previous output, e.g.,
    - Type 2 is on average 3.89 larger than Type 1 on the Borg scale (need to know that contr.treatment was used!). 95%-CI: $(2.9, 4.9)$.
    - Type 3 is on average 2.22 larger than Type 1 on the Borg scale. 95%-CI: $(1.2, 3.2)$.
    - etc.
  - Intercept: **Expected value** for Type 1 (randomly pick a new subject).
  - Subjects have a standard deviation of $\hat{\sigma}_\beta = 1.33$, 95%-CI: $(0.7, 2.3)$.
  - The error standard deviation is $\hat{\sigma} = 1.1$, 95%-CI $(0.8, 1.4)$.

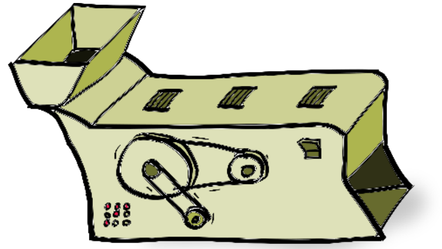
Example: Stools – Alternative Approach

- We could also interpret subject as a **fixed** block factor and do the analysis with `aov`.

```
> fit2 <- aov(effort ~ Type + Subject, data = ergoStool)
> summary(fit2)
            Df Sum Sq Mean Sq F value   Pr(>F)    
Type         3  81.19  27.065  22.356 3.93e-07 ***
Subject      8  66.50   8.313   6.866 0.000106 ***
Residuals   24  29.06   1.211                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> coef(fit2)
(Intercept)      TypeT2      TypeT3      TypeT4    Subject5    Subject4    Subject9 
  6.5555556   3.8888889   2.2222222   0.6666667   0.2500000   1.0000000   1.7500000 
   Subject6    Subject3    Subject7    Subject1    Subject2 
  2.0000000   2.5000000   2.5000000   4.0000000   4.0000000 
```

- Treatment effects are the same (be careful with the meaning of intercept).
  - here: The intercept corresponds to **reference treatment**, **reference subject**.
  - before: The intercept **corresponded to reference treatment**, **expected value over all subjects**.
- Even the p-value of the F-test for treatment is the same. Reason: We don’t model a subject specific treatment effect.
- Of course there is **no** variance component of subject.

Examples: Machines



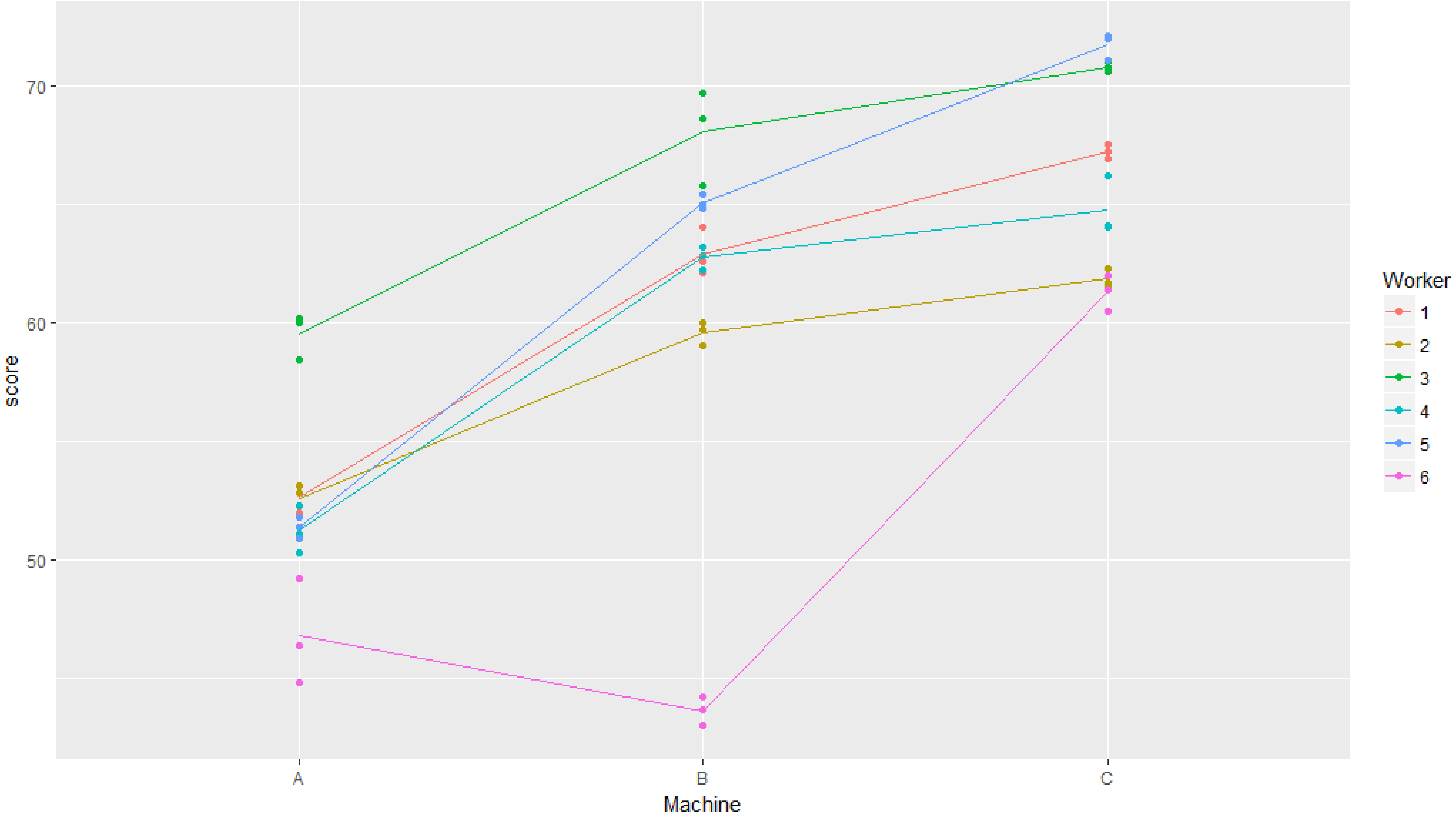
- Dataset `Machines` from R-package `nlme`.

- As stated in the help file:

  *Data on an experiment to compare **three brands of machines** used in an industrial process are presented in Milliken and Johnson (p. 285, 1992). **Six workers** were chosen **randomly** among the employees of a factory to operate **each machine three times**. The **response** is an overall **productivity score** taking into account the number and quality of components produced.*

- Overview of data
  - 3 different **machines** (A, B, C)
  - 6 different **workers** (randomly selected)
  - 3 measurements per combination of machine and worker: **Productivity score**

Examples: Machines - Visualization



Examples: Machines - Model

- We analyze the data with the following **mixed effects** model:

$$\underbrace{Y_{ijk}}_{\text{productivity score}} = \mu + \underbrace{\alpha_i}_{\text{fixed effect of } \textbf{machine}} + \overbrace{\underbrace{\beta_j}_{\text{random effect of } \textbf{worker}}}^{N(0,\sigma_\beta^2)} + \overbrace{\underbrace{(\alpha\beta)_{ij}}_{\text{random interaction effect } \textbf{between machine and worker}}}^{N(0,\sigma_{\alpha\beta}^2)} + \overbrace{\underbrace{\epsilon_{ijk}}_{\text{error term}}}^{N(0,\sigma^2)}$$

- We assume the unrestricted model for the interaction (as this is what is implemented in `lmer`).

- We fit the model using

```
> fit <- lmer(score ~ Machine + (1 | Worker) + (1 | Worker:Machine), data = Machines)
```

`Machine`: fixed effect of **machine**

`(1 | Worker)`: random effect per **worker**

`(1 | Worker:Machine)`: random interaction **between worker and machine**

Examples: Machines - Output

- The standard output is

```
> summary(fit)
Linear mixed model fit by REML t-tests use Satterthwaite approximations to degrees of freedom [
merModLmerTest]
Formula: score ~ Machine + (1 | Worker) + (1 | Worker:Machine)
   Data: Machines

REML criterion at convergence: 215.7

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.26959 -0.54847 -0.01071  0.43937  2.54006

Random effects:
 Groups         Name        Variance Std.Dev.
 Worker:Machine (Intercept) 13.9095  3.7295
 Worker         (Intercept) 22.8584  4.7811
 Residual                    0.9246  0.9616
Number of obs: 54, groups:  Worker:Machine, 18; Worker, 6

Fixed effects:
            Estimate Std. Error     df t value Pr(>|t|)
(Intercept)   52.356      2.486  8.522  21.062 1.20e-08 ***
MachineB       7.967      2.177 10.000   3.660  0.00439 **
MachineC      13.917      2.177 10.000   6.393 7.91e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

$\hat{\sigma}_{\alpha\beta}$

$\hat{\sigma}_{\beta}$

$\hat{\sigma}$

$\hat{\mu}$

$\hat{\alpha}_2$

$\hat{\alpha}_3$

Coefficients in terms of the “coded” variables. Need to know encoding scheme for interpretation.

Examples: Machines - Output

- We can perform the **global** F-test for **machine** by calling anova:

```
> anova(fit)
Analysis of Variance Table of type III  with  Satterthwaite
approximation for degrees of freedom
        Sum Sq Mean Sq NumDF DenDF F.value    Pr(>F)
Machine 38.051  19.025     2    10  20.576 0.0002855 ***
```

- We can also test the **variance component** of **workers** and the **interaction** and calculate the **confidence intervals** using

```
> rand(fit)
Analysis of Random effects Table:
               Chi.sq Chi.DF p.value
Worker           5.57      1    0.02 *
Worker:Machine  71.19      1  <2e-16 ***
> confint(fit, oldNames = FALSE)
Computing profile confidence intervals ...
                                 2.5 %    97.5 %
sd_(Intercept)|Worker:Machine  2.3528037  5.431503
sd_(Intercept)|Worker          1.9514581  9.410584
sigma                          0.7759507  1.234966
(Intercept)                   47.3951611 57.315949
MachineB                       3.7380904 12.195243
MachineC                       9.6880904 18.145243
```

Examples: Machines - Interpretation

Interpretation of previous outputs:

- Machine is highly significant (p-value from global F-test).
- Machine effects can be read off from the fixed effects part of the previous output, e.g.,
  - machine B is on average 7.97 larger than machine A (need to know that contr.treatment was used!). 95%-CI: $(3.7, 12.2)$
  - etc.
- Machine effects are **population averages**! Meaning: They have to be interpreted as the average over all possible workers.
- Workers have a standard deviation of $\hat{\sigma}_\beta = 4.78$, 95%-CI: $(2.0, 9.4)$.
- The interaction has a standard deviation of $\hat{\sigma}_{\alpha\beta} = 3.73$, 95%-CI: $(2.4, 5.4)$.
- The error standard deviation is $\hat{\sigma} = 0.96$, 95%-CI $(0.8, 1.2)$.

What if we Use a Purely Fixed Effects Model?

- We fit the model with `aov` and get

```
> fit2 <- aov(score ~ Machine * Worker, data = Machines)
> summary(fit2)
               Df Sum Sq Mean Sq F value Pr(>F)
Machine         2 1755.3   877.6  949.17 <2e-16 ***
Worker          5 1241.9   248.4  268.62 <2e-16 ***
Machine:Worker 10  426.5    42.7   46.13 <2e-16 ***
Residuals      36   33.3     0.9
```

- Everything is **much more significant**! Why?

What if We Use a Purely Fixed Effects Model?

- The **mixed effects model** assumes that there is a **population average of the machine effect** (the α_i's).
- It means: What is the machine effect **averaged over the whole population of workers**?
- What we observe in our data is a "contaminated" version (because every **worker** has its **own individual deviation** due to the **random interaction term**).
- Basically, we have 6 observations of the treatment effect and try to estimate the **population average** with them.
- The fixed effects model makes a statement about the **average machine effect of the observed 6 workers**, **not** about the population average! This is **easier**, hence the p-values are **smaller**!

Fitting Mixed Effects Models with aov

OLD

- The function `aov` can be used to fit “easy” mixed models by using an additional `Error()` term.

```
> fit3 <- aov(score ~ Machine + Error(Worker + Machine:Worker), data = Machines)
> summary(fit3)

Error: Worker
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  5   1242   248.4

Error: Worker:Machine
          Df Sum Sq Mean Sq F value   Pr(>F)
Machine    2 1755.3   877.6   20.58 0.000286 ***
Residuals 10  426.5    42.7
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Error: Within
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals 36  33.29  0.9246
```

- We simply **put all the random effects** in `Error()`.

Fitting Mixed Effects Models with aov

- In this example, the p-values coincide with `lmer`.
- In an unbalanced data-set, `aov` can only do Type I sums of squares, `drop1` is no more possible.
- `lmer` is much more flexible in general.
- However, (too) many **theoretical aspects are still unknown**, see for example http://bbolker.github.io/mixedmodels-misc/glmmFAQ.html.
- Nevertheless, mixed models are **extremely popular** in many applied areas.

Back to the Cheese Rating Example

- See the corresponding R-File.