METHODOLOGICAL WORKSHOP 10/03/2025

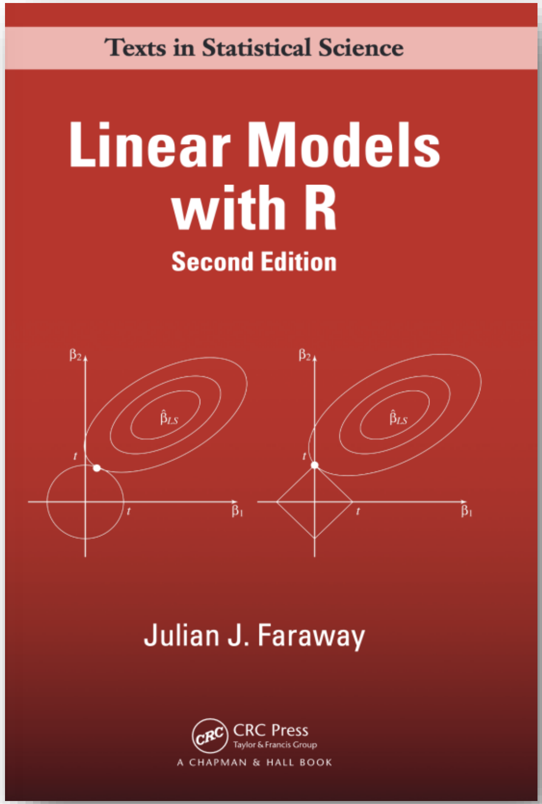
LINEAR MIXED MODELS

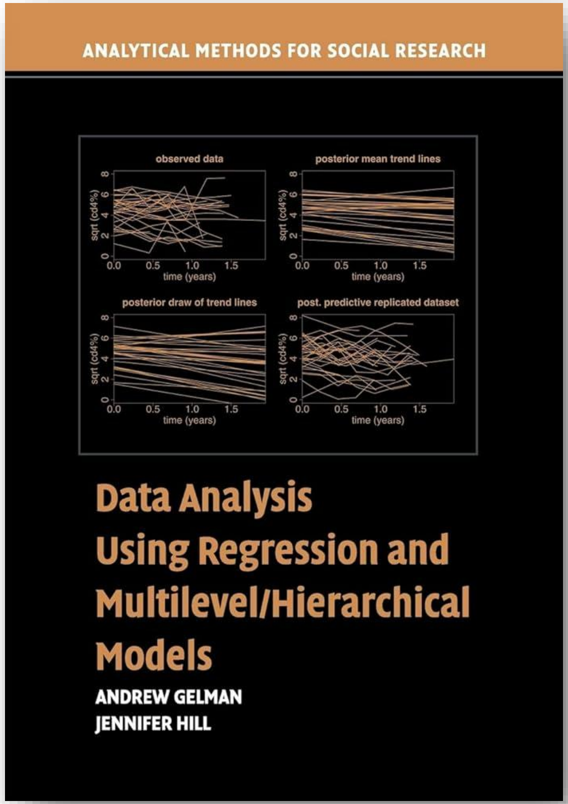
OR MULTILEVEL MODELS OR HIERARCHICAL MODELS OR MIXED MODELS OR NESTED DATA MODELS OR RANDOM-EFFECTS MODELS

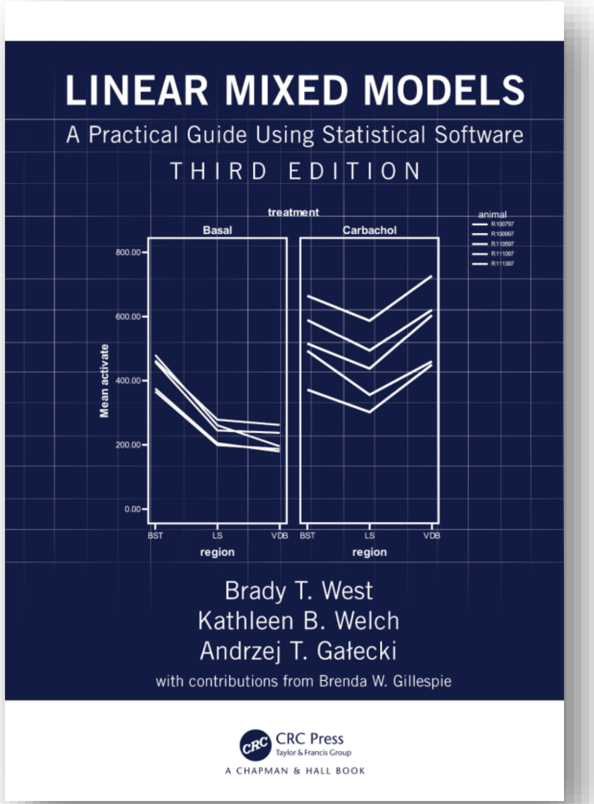
Martin Le Guennec (PhD student)

Disclaimer

I do not pretend to be a statistician! Just a PhD student who tries to understand the tools he uses… So I have read some books







Linear models: The concept

*"Linear regression is a method that summarizes how the **average values of a numerical outcome variable** vary over subpopulations defined by **linear functions or predictors**."*

Gelman & Hill, 2007, *Ch. 3, p. 31*

Linear models: The concept

*"Linear regression is a method that summarizes how the **average values of a numerical outcome variable** vary over subpopulations defined by **linear functions or predictors**."*

Gelman & Hill, 2007, *Ch. 3, p. 31*

The theoretical model

$$
\begin{aligned}
Y &= f(X_1, \ldots, X_k) + \varepsilon, \\
&= \beta_0 + \beta_1 X_1 + \cdots + \beta_2 X_2 + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, \sigma)
\end{aligned}
$$

Linear models: The concept

*"Linear regression is a method that summarizes how the **average values of a numerical outcome variable** vary over subpopulations defined by **linear functions or predictors**."*

Gelman & Hill, 2007, *Ch. 3, p. 31*

The theoretical model

$$\begin{aligned} Y &= f(X_1, \dots, X_k) + \varepsilon, \\ &= \beta_0 + \beta_1 X_1 + \cdots + \beta_2 X_2 + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, \sigma) \end{aligned}$$

The matrix notation is useful for further theoretical development:

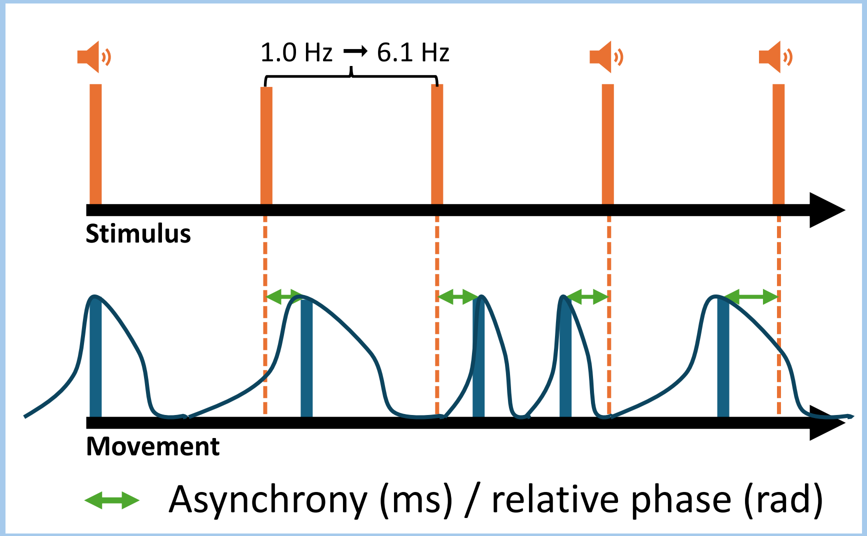
$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

or explicitly:

$$\begin{pmatrix} y_0 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & \dots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_n \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

Linear models: The concept

Illustration: Sensorimotor Synchronization



Sample: 15 Indian and 13 French participants

Task:

- « Tap exactly at the same time as the metronome »
- Increasing frequency of the metronome
- 7 repetition of the task for each participant

Linear models: The concept

Illustration: Sensorimotor Synchronization

Imagine we can describe the asynchrony as:

$$\phi = -100 + 50\omega + \varepsilon,$$

and we have the following values of ϕ:

	rate	asynchrony
1	1.0	-66.81
2	1.3	-41.91
3	1.6	26.76
4	1.9	-2.88
5	2.2	13.88
6	2.5	76.45
7	2.8	53.83
8	3.1	17.05
9	3.4	49.39
10	3.7	71.63

Since we know the theoretical relationship between rate and asynchrony, we can determine ε from the ϕ values. Let's write the developped form of:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

Linear models: The concept

Illustration: Sensorimotor Synchronization

	rate	asynchrony
1	1.0	-66.81
2	1.3	-41.91
3	1.6	26.76
4	1.9	-2.88
5	2.2	13.88
6	2.5	76.45
7	2.8	53.83
8	3.1	17.05
9	3.4	49.39
10	3.7	71.63

The structure of the relationship is as follows:

$$\begin{bmatrix} \phi_0 \\ \vdots \\ \phi_9 \end{bmatrix} =$$

and if we replace with the actual values:

$$\begin{bmatrix} -66.81 \\ -41.91 \\ 26.76 \\ -2.88 \\ 13.48 \\ 76.45 \\ 53.83 \\ 17.05 \\ 49.39 \\ 71.63 \end{bmatrix} =$$

Linear models: The concept

Illustration: Sensorimotor Synchronization

The structure of the relationship is as follows:

$$\begin{bmatrix}\phi_0\\ \vdots \\ \phi_9\end{bmatrix} = \begin{bmatrix}X_{0,0} & X_{1,0}\\ \vdots & \vdots \\ X_{0,9} & X_{1,9}\end{bmatrix}\begin{bmatrix}\beta_0\\ \beta_1\end{bmatrix}$$

	rate	asynchrony
1	1.0	-66.81
2	1.3	-41.91
3	1.6	26.76
4	1.9	-2.88
5	2.2	13.88
6	2.5	76.45
7	2.8	53.83
8	3.1	17.05
9	3.4	49.39
10	3.7	71.63

and if we replace with the actual values:

$$\begin{bmatrix}-66.81\\ -41.91\\ 26.76\\ -2.88\\ 13.48\\ 76.45\\ 53.83\\ 17.05\\ 49.39\\ 71.63\end{bmatrix} = \begin{bmatrix}1 & 1.0\\ 1 & 1.3\\ 1 & 1.6\\ 1 & 1.9\\ 1 & 2.2\\ 1 & 2.5\\ 1 & 2.8\\ 1 & 3.1\\ 1 & 3.4\\ 1 & 3.7\end{bmatrix}\begin{bmatrix}-100\\ 50\end{bmatrix}$$

$$\begin{bmatrix}a_{00} & a_{01} & a_{02}\\ a_{10} & a_{11} & a_{12}\\ a_{20} & a_{21} & a_{22}\end{bmatrix} \times \begin{bmatrix}b_{00} & b_{01} & b_{02}\\ b_{10} & b_{11} & b_{12}\\ b_{20} & b_{21} & b_{22}\end{bmatrix} = \begin{bmatrix}c_{00} & c_{01} & c_{02}\\ c_{10} & c_{11} & c_{12}\\ c_{20} & c_{21} & c_{22}\end{bmatrix}$$

$$a_{00} \times b_{00} + a_{01} \times b_{10} + a_{02} \times b_{20} = c_{00}$$

Linear models: The concept

Illustration: Sensorimotor Synchronization

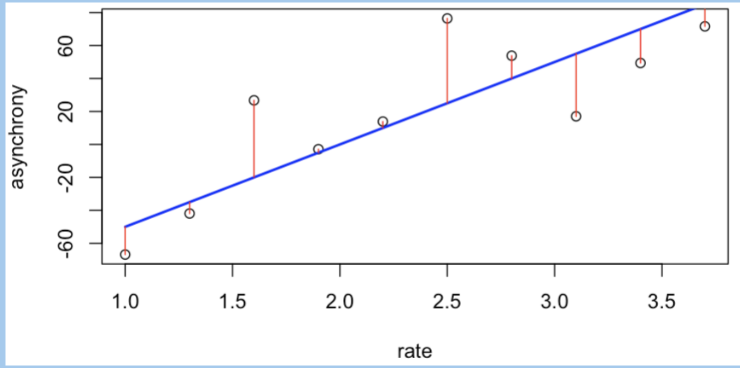
	rate	asynchrony
1	1.0	-66.81
2	1.3	-41.91
3	1.6	26.76
4	1.9	-2.88
5	2.2	13.88
6	2.5	76.45
7	2.8	53.83
8	3.1	17.05
9	3.4	49.39
10	3.7	71.63

The structure of the relationship is as follows:

$$\begin{bmatrix} \phi_0 \\ \vdots \\ \phi_9 \end{bmatrix} = \begin{bmatrix} X_{0,0} & X_{1,0} \\ \vdots & \vdots \\ X_{0,9} & X_{1,9} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_0 \\ \vdots \\ \varepsilon_9 \end{bmatrix},$$

and if we replace with the actual values:

$$\begin{bmatrix} -66.81 \\ -41.91 \\ 26.76 \\ -2.88 \\ 13.48 \\ 76.45 \\ 53.83 \\ 17.05 \\ 49.39 \\ 71.63 \end{bmatrix} = \begin{bmatrix} 1 & 1.0 \\ 1 & 1.3 \\ 1 & 1.6 \\ 1 & 1.9 \\ 1 & 2.2 \\ 1 & 2.5 \\ 1 & 2.8 \\ 1 & 3.1 \\ 1 & 3.4 \\ 1 & 3.7 \end{bmatrix} \begin{bmatrix} -100 \\ 50 \end{bmatrix} + \begin{bmatrix} -16.81 \\ -6.91 \\ 46.76 \\ 2.12 \\ 3.88 \\ 51.45 \\ 13.82 \\ -37.95 \\ -20.61 \\ -13.37 \end{bmatrix}.$$



Linear models: Estimation

We can't directly observe the true coefficients β or the errors ε

→ We have to **estimate** them

- β refers to the theoretical values, $\hat{\beta}$ refers to the coefficients estimated from the data
- ε refers to the theoretical errors, $\hat{\varepsilon}$ refers to the residuals estimated from the data

Linear models: Estimation

We can't directly observe the true coefficients β or the errors ε

→ We have to **estimate** them

- β refers to the theoretical values, $\hat{\beta}$ refers to the coefficients estimated from the data
- ε refers to the theoretical errors, $\hat{\varepsilon}$ refers to the residuals estimated from the data

Gauss-Markov theorem

Suppose $\mathbb{E}[\varepsilon] = 0$ and $\mathrm{Var}(\varepsilon) = \sigma^2 I$. Suppose also that the structural part of the model $\mathbb{E}[\mathrm{Y}] = X\beta$ is correct. Let $\boldsymbol{\psi} = c^\intercal\beta$ be an estimable function; then, in the class of all unbiased linear estimates of $\boldsymbol{\psi}$, $\widehat{\boldsymbol{\psi}} = c^\intercal\hat{\beta}$, where $\hat{\beta}$ is the ordinary least squares estimator, has the minimum variance and is unique.

Farway, 2015, *Ch. 2, p. 22*

Linear models: Estimation

We can't directly observe the true coefficients β or the errors ε

→ We have to **estimate** them

- β refers to the theoretical values, $\hat{\beta}$ refers to the coefficients estimated from the data
- ε refers to the theoretical errors, $\hat{\varepsilon}$ refers to the residuals estimated from the data

Gauss-Markov theorem

Suppose $\mathbb{E}[\varepsilon] = 0$ and $\mathrm{Var}(\varepsilon) = \sigma^2 I$. Suppose also that the structural part of the model $\mathbb{E}[\mathrm{Y}] = X\beta$ is correct. Let $\boldsymbol{\psi} = c^\intercal \beta$ be an estimable function; then, in the class of all unbiased linear estimates of $\boldsymbol{\psi}$, $\widehat{\boldsymbol{\psi}} = c^\intercal \hat{\beta}$, where $\hat{\beta}$ is the ordinary least squares estimator, has the minimum variance and is unique.

Farway, 2015, *Ch. 2, p. 22*

Assumptions of the Gauss-Markov theorem

- Linear relationship
- Uncorrelated residuals
- Homoscedasticity (equal variance) of residuals
- Mean value of the residuals = 0

Ordinary least squares (OLS) estimates are the best linear unbiased estimates (BLUE)

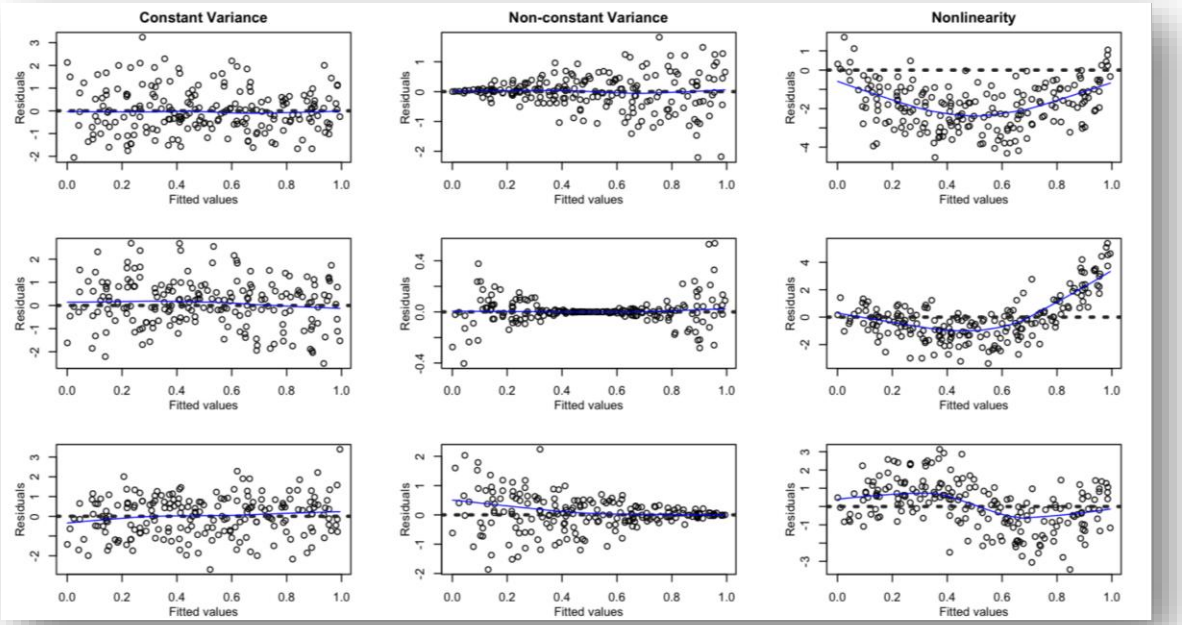
Linear models: Estimation

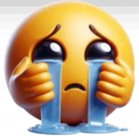
When fitting a linear regression, you have to verify these asumptions !!!

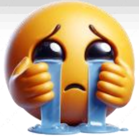
Relationship linearity & Homoscedasticity

Verify with a fitted vs. residual plot

```
> model <- lm(x ~ y)
> plot(model, which = 1)
```







Non-linearity = mean structure conceptually wrong

Heteroscedasticity = something about variability is not modelled

Can be accounted for:

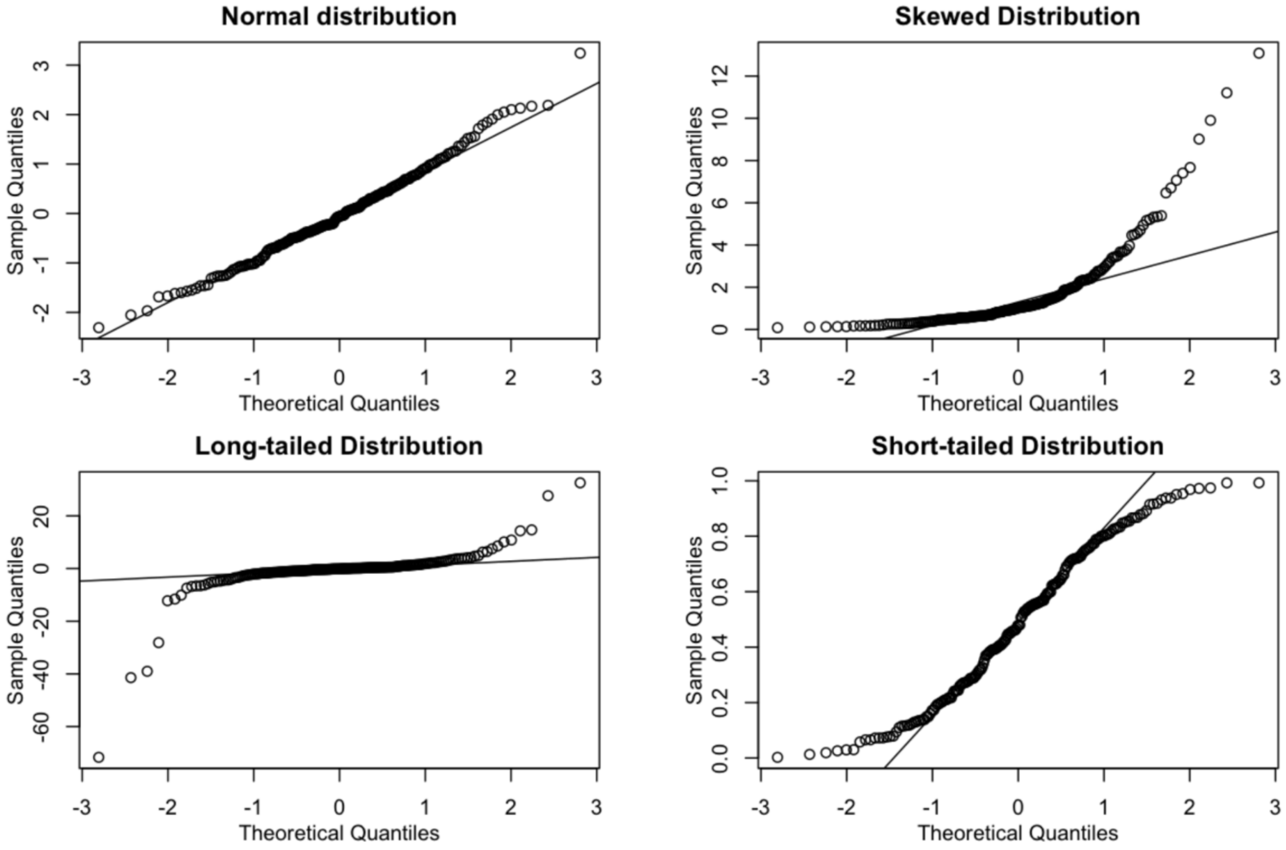
- Heteroscedasticity Consistent Covariance Matrix
- Generalized Least Squares
- Variance-stabilizing transformations of the dependent variable

Long & Ervin, 2000, *Am Stat*; Faraway, 2015

Linear models: Estimation

What about normality!?

- Not necessary for an unbiased estimation
- Essential for making valid inferences (assumptions from the tests)



Linear models: Estimation

Let's illustrate why we talk about estimation

```
> set.seed(123)

> # Simulate the 7 task for each participant
> rate <- seq(from = 1, to = 3.7, by = 0.3)
> rate <- rep(rate, 7)

> fit <- 50 * rate – 100  # Theoretical relationship
> noise <- rnorm(n = length(rate), mean = 0, sd = 30)
> asynchrony <- fit + noise  # Measured asynchrony

> # Estimate parameters using least squares
> model <- lm(asynchrony ~ rate)
> summary(model)
```

Theoretical

$y = -100.00 + 50.00x$

Linear models: Estimation

Let’s illustrate why we talk about estimation

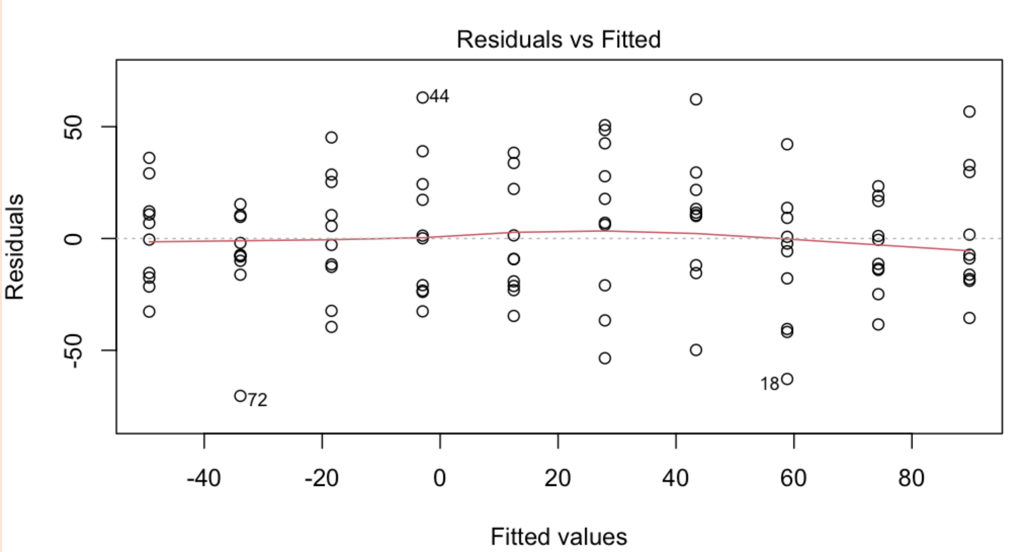
```
> set.seed(123)

> # Simulate the 7 task for each participant
> rate <- seq(from = 1, to = 3.7, by = 0.3)
> rate <- rep(rate, 7)

> fit <- 50 * rate – 100  # Theoretical relationship
> noise <- rnorm(n = length(rate), mean = 0, sd = 30)
> asynchrony <- fit + noise  # Measured asynchrony

> # Estimate parameters using least squares
> model <- lm(asynchrony ~ rate)
> summary(model)

Call:
lm(formula = asynchrony ~ rate)

Residuals:
    Min      1Q  Median      3Q     Max
-61.990 -17.471  -0.517  16.941  63.320

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -100.218      9.516  -10.53 6.29e-16 ***
rate          51.035      3.802   13.42  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 27.41 on 68 degrees of freedom
Multiple R-squared:  0.726,	Adjusted R-squared:  0.722
F-statistic: 180.2 on 1 and 68 DF,  p-value: < 2.2e-16
```

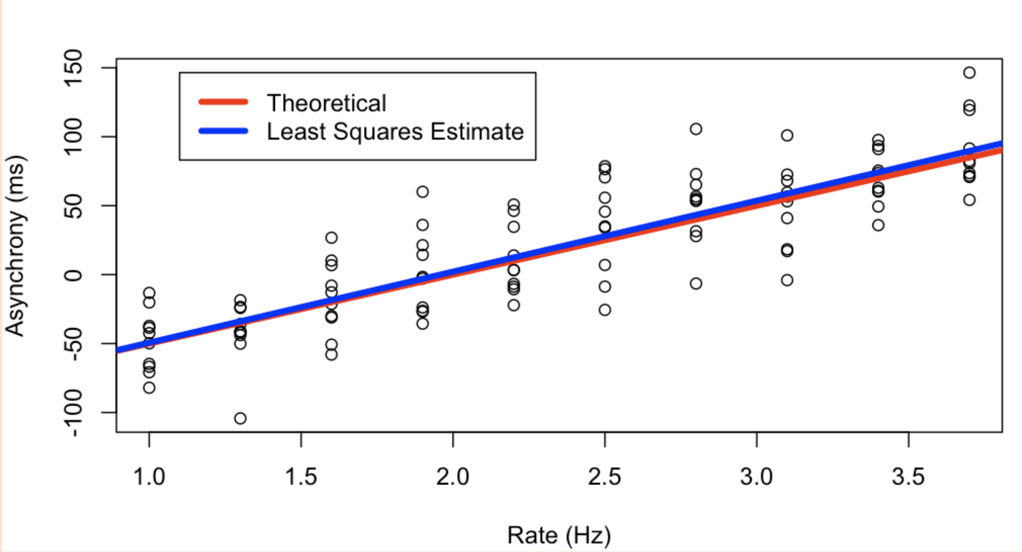
Theoretical

➡ $y = -100.00 + 50.00x$

OLS

➡ $y = -100.22 + 51.03x$





Linear models: Estimation

Quick detour on inferential testing

> summary(model)

```
Call:
lm(formula = asynchrony ~ rate)

Residuals:
    Min      1Q  Median      3Q     Max 
-61.990 -17.471  -0.517  16.941  63.320 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -100.218      9.516  -10.53 6.29e-16 ***
rate          51.035      3.802   13.42  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 27.41 on 68 degrees of freedom
Multiple R-squared:  0.726,	Adjusted R-squared:  0.722 
F-statistic: 180.2 on 1 and 68 DF,  p-value: < 2.2e-16
```

For every coefficient, we test:

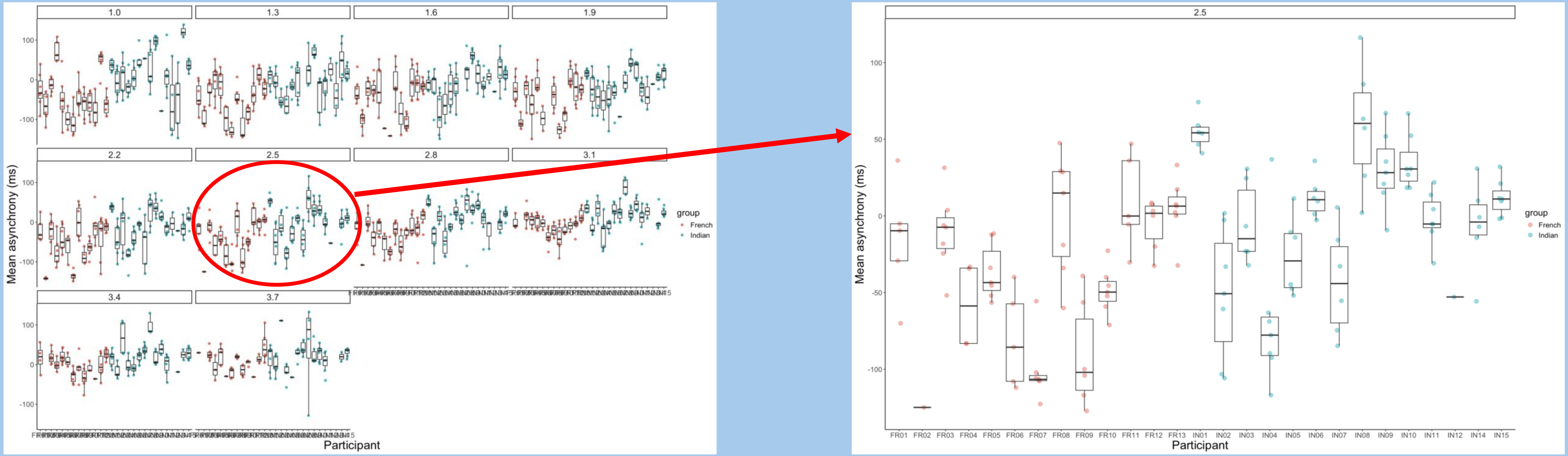
$$H_0: \beta = 0,$$

using a t-test with the t-value being:

$$t = \frac{\text{Estimate}}{\text{SE}}.$$

The number of df is the number of observation – the number of parameters

Linear mixed models: Illustration



- Correlated measures
  - Multiple trials for a participant at each rate
- Unbalanced design
  - Mean asynchrony only makes sense when there is stable synchronization

Assumptions of the Gauss-Markov theorem are not respected: OLS estimates are not the BLUE

Leads to various estimation problems

Montgommery, 2017

```
> rstatix::anova_test(
+   data = df_async_by_task,
+   dv = mean_async,
+   wid = subject,
+   within = frequency,
+   between = group
+ )
Error in `spread()`:
! Each row of output must be identified by a unique combination of keys.
i Keys are shared for 1458 rows
```

Linear mixed models: Illustration

Linear mixed models solve those problems:

- Intra-individual variation (source of correlation) is included in the model
- Unbalanced and incomplete designs do not yield estimation problems

Linear mixed models: Illustration

Linear mixed models solve those problems:

- Intra-individual variation (source of correlation) is included in the model
- Unbalanced and incomplete designs do not yield estimation problems

Example focusing on one rate: testing group difference at 1.0 Hz

Let's consider two different approaches to estimate the group averages

Calculate the overall average *« Complete pooling »* Completely ignores variation between participants	**Calculate an average for each participant** *« No pooling »* Overfits the data within each participants Estimates may be influenced by outliers for participants with few measurements

Linear mixed models: Illustration

Linear mixed models solve those problems:

- Intra-individual variation (source of correlation) is included in the model
- Unbalanced and incomplete designs do not yield estimation problems

Example focusing on one rate: testing group difference at 1.0 Hz

Let’s consider two different approaches to estimate the group averages

Calculate the overall average

« Complete pooling »

Completely ignores variation between participants

Calculate an average for each participant

« No pooling »

Overfits the data within each participants

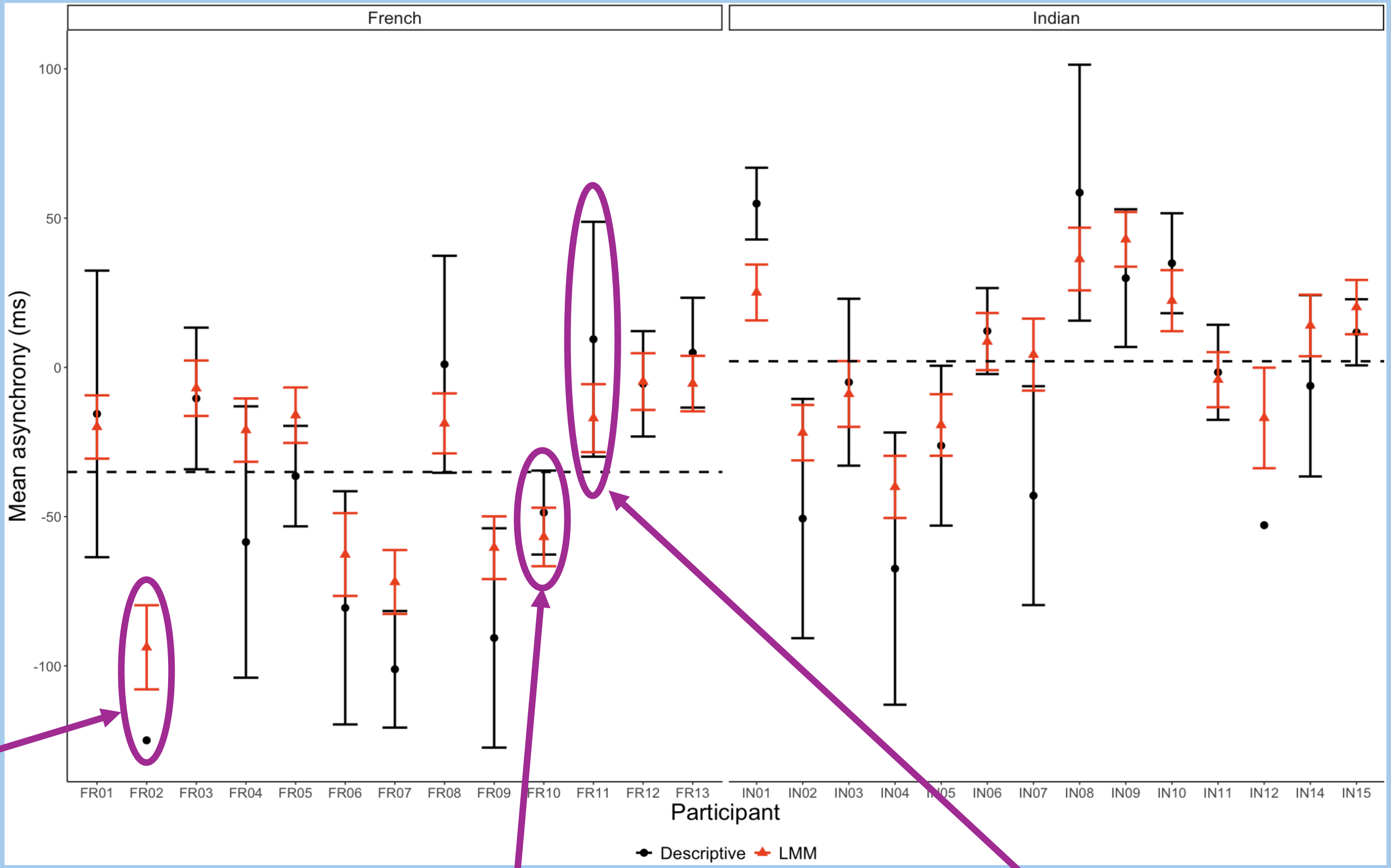
Estimates may be influenced by outliers for participants with few measurements

Linear mixed models estimates are a compromise between these extremes:

Participants’ averages are pulled toward the group average

Participants with few or very variable observations are more pulled

Linear mixed models: Illustration



- Summary stats
- LMM stats

1 value, mean value strongly attracted to group average

Low variability, mean value not pulled toward group mean

High variability, mean value pulled toward group mean

Linear mixed models

When should I use linear mixed models over linear models ?

“Our advice is to ***always use multilevel modeling*** *[…] Classical regressions can typically be identified with particular special cases of multilevel models with hierarchical variance parameters set to zero or infinity—these are the complete pooling and no pooling models.”*

Gelman & Hill, 2007, *Ch. 11*

Linear mixed models

We are going to discuss technical details to introduce the two estimation methods

If this is too technical, just remember that both methods estimate the *fixed-effect parameters* and the *variance components* (random-effects and residuals) of the model.

However:

- **ML estimates are biased** because there is a loss of df in the estimation process not accounted for
- **REML estimates corrects this bias** by taking into account the loss of df

West, Welch & Gałecki, 2022

Linear mixed models: The concept

Matrix notation for a linear mixed model

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\varepsilon}$$

- $\boldsymbol{Y}$: vector of observations with $\mathbb{E}[\boldsymbol{Y}] = \boldsymbol{X\beta}$
- $\boldsymbol{X}$: design matrix for fixed effects (group level)
- $\boldsymbol{Z}$: design matrix for random effects (participant level)
- $\boldsymbol{u}$: random effects with $\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{D})$
- $\boldsymbol{\varepsilon}$: residuals with $\boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R})$

West, Welch & Gałecki, 2022

Linear mixed models: The concept

Matrix notation for a linear mixed model

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \color{red}{\boldsymbol{Zu}} + \boldsymbol{\varepsilon}$$

- $\boldsymbol{Y}$: vector of observations with $\mathbb{E}[\boldsymbol{Y}] = \boldsymbol{X\beta}$
- $\boldsymbol{X}$: design matrix for fixed effects (group level)
- $\boldsymbol{Z}$: design matrix for random effects (participant level)
- $\boldsymbol{u}$: random effects with $\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{D})$
- $\boldsymbol{\varepsilon}$: residuals with $\boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R})$

The elements of $\boldsymbol{D}$ and $\boldsymbol{R}$ are functions of a small set of covariance parameters stored in the vectors $\boldsymbol{\theta}_D$ and $\boldsymbol{\theta}_R$, combined in a vector $\boldsymbol{\theta}$

$$\boldsymbol{D} = \mathrm{Var}(\boldsymbol{u}) = \begin{pmatrix} \mathrm{Var}(u_1) & \mathrm{Cov}(u_1, u_2) & \cdots & \mathrm{Cov}(u_1, u_q) \\ \mathrm{Cov}(u_1, u_2) & \mathrm{Var}(u_2) & \cdots & \mathrm{Cov}(u_2, u_q) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(u_1, u_q) & \mathrm{Cov}(u_2, u_q) & \cdots & \mathrm{Var}(u_q) \end{pmatrix} \Rightarrow \boldsymbol{\theta}_D = \begin{pmatrix} \mathrm{Var}(u_1) \\ \mathrm{Var}(u_2) \\ \vdots \\ \mathrm{Var}(u_q) \\ \mathrm{Cov}(u_1, u_2) \\ \vdots \end{pmatrix}$$

West, Welch & Gałecki, 2022

Linear mixed models: The concept

Marginal model

The marginal model of a linear mixed model does not contain the random-effect structure:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon},$$

with $\boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{V})$ and:

$$\boldsymbol{V} = \boldsymbol{ZDZ}' + \boldsymbol{R},$$

so the elements of $\boldsymbol{V}$ are functions of the elements of $\boldsymbol{\theta}$.

This defines the marginal distribution of $\boldsymbol{Y}$:

$$\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{X\beta}, \boldsymbol{ZDZ}' + \boldsymbol{R})$$

Linear mixed models: Estimation

Maximum Likelihood (ML) estimation

- Construct a likelihood function ℓ_{ML} of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ using the marginal distribution of $\boldsymbol{Y}$

Linear mixed models: Estimation

Maximum Likelihood (ML) estimation

- Construct a likelihood function ℓ_{ML} of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ using the marginal distribution of $\boldsymbol{Y}$
- Obtain BLUE of $\widehat{\boldsymbol{\beta}}$ for a given value of $\boldsymbol{\theta}$

$$\ell_{ML}(\boldsymbol{\theta}) = -0.5n \times \ln(2\pi) - 0.5 \sum_i \ln(\det(\boldsymbol{V})) - 0.5 \sum_i \boldsymbol{r}' \boldsymbol{V}^{-1} \boldsymbol{r}$$

Linear mixed models: Estimation

Maximum Likelihood (ML) estimation

- Construct a likelihood function ℓ_{ML} of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ using the marginal distribution of $\boldsymbol{Y}$
- Obtain BLUE of $\widehat{\boldsymbol{\beta}}$ for a given value of $\boldsymbol{\theta}$
- Obtain optimal estimates $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ using iterative computations to maximize $\ell_{ML}(\boldsymbol{\theta})$, until reaching convergence

$$\ell_{ML}(\boldsymbol{\theta}) = -0.5n \times \ln(2\pi) - 0.5 \sum_i \ln(\det(\boldsymbol{V})) - 0.5 \sum_i \boldsymbol{r}' \, \boldsymbol{V}^{-1} \boldsymbol{r}$$

Linear mixed models: Estimation

Maximum Likelihood (ML) estimation

- Construct a likelihood function ℓ_{ML} of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ using the marginal distribution of $\boldsymbol{Y}$
- Obtain BLUE of $\widehat{\boldsymbol{\beta}}$ for a given value of $\boldsymbol{\theta}$
- Obtain optimal estimates $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ using iterative computations to maximize $\ell_{ML}(\boldsymbol{\theta})$, until reaching convergence
- Calculate again $\widehat{\boldsymbol{\beta}}$ using $\widehat{\boldsymbol{\theta}}$: those are the ML fixed effects and covariance parameters

$$\ell_{ML}(\boldsymbol{\theta}) = -0.5n \times \ln(2\pi) - 0.5 \sum_i \ln(\det(\boldsymbol{V})) - 0.5 \sum_i \boldsymbol{r}'\, \boldsymbol{V}^{-1} \boldsymbol{r}$$

Linear mixed models: Estimation

Maximum Likelihood (ML) estimation

- Construct a likelihood function ℓ_{ML} of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ using the marginal distribution of $\boldsymbol{Y}$
- Obtain BLUE of $\widehat{\boldsymbol{\beta}}$ for a given value of $\boldsymbol{\theta}$
- Obtain optimal estimates $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ using iterative computations to maximize $\ell_{ML}(\boldsymbol{\theta})$, until reaching convergence
- Calculate again $\widehat{\boldsymbol{\beta}}$ using $\widehat{\boldsymbol{\theta}}$: those are the ML fixed effects and covariance parameters

Restricted Maximum Likelihood (ML) estimation

The ML estimates of $\boldsymbol{\theta}$ are biased, loss of degrees of freedom from estimating parameters $\boldsymbol{\beta}$ is not taken into account. REML corrects that and produces unbiased estimates

$$\ell_{ML}(\boldsymbol{\theta}) = -0.5n \times \ln(2\pi) - 0.5\sum_i \ln(\det(\boldsymbol{V})) - 0.5\sum_i \boldsymbol{r}'\,\boldsymbol{V}^{-1}\boldsymbol{r}$$

$$\ell_{REML}(\boldsymbol{\theta}) = -0.5(n-p) \times \ln(2\pi) - 0.5\sum_i \ln(\det(\boldsymbol{V})) - 0.5\sum_i \boldsymbol{r}'\,\boldsymbol{V}^{-1}\boldsymbol{r} - 0.5\sum_i \ln(\det(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}))$$

Linear mixed models: Estimation

REML estimation is preferred over ML estimation

However, ML is usefull to chose the fixed-effect structure:

- We can't compare REML likelihood from models with different fixed-effect structures (different transformed data space)
- ML likelihood is comparable across models with different fixed-effects structures

Use REML when comparing models with the same fixed-effect structure but with different random-effect structures

Useful R libraries

- *lme4* — Bates et al., 2015, *J Stat Soft*
  - **Purpose:** Fit linear mixed models using ML or REML
  - **Outcome:** fixed-effect and random-effect coefficients
- *lmerTest* — Kuznetsova, Brockhoff & Christensen, 2017, *J Stat Soft*
  - **Purpose:** Inference testing on fixed-effects parameters obtained with *lme4*
  - **Outcome:** degrees of freedom, *t* statistic, *p*-value

Remember to cite correctly the packages your are using in your analyses! Not only does it help reproduce your work, it also supports all the work done by the authors.

In R:

```
> citation("lme4")
```

Linear mixed models: Formula in R

To fit a linear mixed model, we use the *lmer* function from the package *lmerTest* (or *lme4*):

Fixed effects

Random effects

```
> lmm_ex <- lmer(
+   formula = y ~ x1 + x2 * x3 + (1 | participant),
+   data = df,
+   REML = TRUE
+ )
```

« Model y *as a function of* x1*,* x2*,* x3 *and the interaction between* x2 *and* x3*, with a random intercept for every participant »*

Bates et al., 2015, *J Stat Soft*

Linear mixed models: Formula in R

To fit a linear mixed model, we use the *lmer* function from the package *lmerTest* (or *lme4*):

Fixed effects   Random effects

```
> lmm_ex <- lmer(
+   formula = y ~ x1 + x2 * x3 + (1 | participant),
+   data = df,
+   REML = TRUE
+ )
```

« Model y *as a function of* x1*,* x2*,* x3 *and the interaction between* x2 *and* x3*, with a random intercept for every participant »*

Bates et al., 2015, *J Stat Soft*

How do I determine my formula?

There is not one single answer to that question:

- **Top-down strategy**: select well-specific mean structure and random effect structure, then reduce the model
- **Step-up strategy**: begin with means-only, then consider adding random effects if it improves the model
- **Maximum structure**: use random intercept and slope for every intra-subject factor
- …

West, Welch & Gałecki, 2022

Linear mixed models: Formula in R

How did I want to analyze my data?

1. ***What do I want to test?***
   - I want to test whether the mean asynchrony differs between group at each metronome rate
   - My model should test the main effect of group and rate, as well as their interaction

Linear mixed models: Formula in R

How did I want to analyze my data?

1. ***What do I want to test?***
   - I want to test whether the mean asynchrony differs between group at each metronome rate
   - My model should test the main effect of group and rate, as well as their interaction
2. ***What is the structure of my data set?***
   - Several trials for every participants, there is one value of mean asynchrony per trial per rate per participant if the participant was synchronized, otherwise no value for the rate of the trial
   - My model should include a random intercept for every participant

Linear mixed models: Formula in R

How did I want to analyze my data?

1. *What do I want to test?*

- I want to test whether the mean asynchrony differs between group at each metronome rate
- My model should test the main effect of group and rate, as well as their interaction

2. *What is the structure of my data set?*

- Several trials for every participants, there is one value of mean asynchrony per trial per rate per participant if the participant was synchronized, otherwise no value for the rate of the trial
- My model should include a random intercept for every participant

3. *Should I only include a random intercept or a random intercept for each rate?*

- I have no theoretical reason to believe that one structure is preferable over the other one, I will compute the AICc weights for each model to choose the best model

  Burnham & Anderson, 2004, *Sociol Methods Res*

Name	AIC (weights)	AICc (weights)	BIC (weights)	R2 (cond.)	R2 (marg.)	ICC	RMSE	Sigma
lmm_ri	14908.0 (0.190)	14908.7 (0.190)	15024.5 (0.190)	0.516	0.217	0.382	35.687	36.245
lmm_rin	14905.1 (0.810)	14905.8 (0.810)	15021.6 (0.810)	0.652	0.234	0.546	28.777	31.309

Table obtained using *compare_performance()* from the *performance* package (Lüdecke et al., 2021, *J Open Source Soft)*

Linear mixed models: Formula in R

To summarize:

- Main effects group and frequency + interaction group-frequency
- Random intercept per participant nested by frequency

```
> lmm_ex <- lmer(
+   mean_async ~ group * frequency + (1 | subject:frequency),
+   data = df,
+   REML = TRUE
+ )
```

Linear mixed models: Formula in R

To summarize:

- Main effects group and frequency + interaction group-frequency
- Random intercept per participant nested by frequency

```
> lmm_ex <- lmer(
+   mean_async ~ group * frequency + (1 | subject:frequency),
+   data = df,
+   REML = TRUE
+ )
```

If my frequency would have been included as a continuous variable, I could have included frequency as a random slope

```
> lmm_ex <- lmer(
+   mean_async ~ group * frequency + (frequency | subject),
+   data = df,
+   REML = TRUE
+ )
```

Linear mixed models: Model selection

Likelihood Ratio Tests (LRTs)

"LRTs are a class of tests that are based on comparing the values of likelihood functions $\mathcal{L}$ for two models defining a hypothesis being tested."

$$-2\log\left(\frac{\mathcal{L}_{mod1}}{\mathcal{L}_{mod2}}\right) \sim \chi^2_{df}$$

West, Welch & Gałecki, 2022, *Ch. 2.6*

Linear mixed models: Model selection

Likelihood Ratio Tests (LRTs)

"LRTs are a class of tests that are based on comparing the values of likelihood functions $\mathcal{L}$ for two models defining a hypothesis being tested."

$$-2\log\left(\frac{\mathcal{L}_{mod1}}{\mathcal{L}_{mod2}}\right) \sim \chi^2_{df}$$

LRT for fixed-effects structure

Estimation method: ML

df: difference in the number of fixed-effect parameters

West, Welch & Gałecki, 2022, *Ch. 2.6*

Linear mixed models: Model selection

Likelihood Ratio Tests (LRTs)

"LRTs are a class of tests that are based on comparing the values of likelihood functions $\mathcal{L}$ for two models defining a hypothesis being tested."

$$-2\log\left(\frac{\mathcal{L}_{mod1}}{\mathcal{L}_{mod2}}\right) \sim \chi^2_{df}$$

LRT for fixed-effects structure	LRT for random-effects structure
Estimation method: ML	**Estimation method**: REML
df: difference in the number of fixed-effect parameters	**df**: depends on covariance parameters (see West et al., 2022)

West, Welch & Gałecki, 2022, *Ch. 2.6*

Linear mixed models: Formula in R

Examples of random effect structures

- Participants differ in their baseline level, but share the same slope for x
  y ~ x + (1 | participant)
- Participants differ in both intercept and slope for x, with correlation between intercept and slope
  y ~ x + (1 + x | participant)
- Participants differ in intercept and main-effect slopes
  y ~ x1 * x2 + (1 + x1 + x2 | participant)
- Participants differ in intercept and main-effect interaction but not slopes
  y ~ x1 * x2 + (1 + x1:x2 | participant)
- Both participants and items contributed independent intercept variation
  y ~ x + (1 | participant) + (1 | item)
- Participants are nested within hospitals, with a random intercept per participant in hospital
  y ~ x (1 | participant/hospital)
- Participant-by-condition random intercepts
  y ~ x + (1 | participant:condition)

Linear mixed models: Formula in R

Examples of random effect structures

Formula	Alternative	Meaning
`(1 | g)`	`1 + (1 | g)`	Random intercept with fixed mean.
`0 + offset(o) + (1 | g)`	`-1 + offset(o) + (1 | g)`	Random intercept with *a priori* means.
`(1 | g1/g2)`	`(1 | g1)+(1 | g1:g2)`	Intercept varying among `g1` and `g2` within `g1`.
`(1 | g1) + (1 | g2)`	`1 + (1 | g1) + (1 | g2).`	Intercept varying among `g1` and `g2`.
`x + (x | g)`	`1 + x + (1 + x | g)`	Correlated random intercept and slope.
`x + (x || g)`	`1 + x + (1 | g) + (0 + x | g)`	Uncorrelated random intercept and slope.

Table 2: Examples of the right-hand-sides of mixed-effects model formulas. The names of grouping factors are denoted `g`, `g1`, and `g2`, and covariates and *a priori* known offsets as `x` and `o`.

From Bates et al., 2015, *J Stat Soft*

Linear mixed models: Illustration

```
> lmm_mean_async <- lmerTest::lmer(
+   mean_async ~ group * frequency + (1 | subject:frequency),
+   data = df_async_by_task
+ )
> summary(lmm_mean_async)

Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: mean_async ~ group * frequency + (1 | subject:frequency)
   Data: df_async_by_task

REML criterion at convergence: 14731.5

Scaled residuals:
    Min      1Q  Median      3Q     Max
-5.4197 -0.5253  0.0020  0.5799  3.4152

Random effects:
 Groups            Name        Variance Std.Dev.
 subject:frequency (Intercept) 1177.9   34.32
 Residual                       980.2   31.31
Number of obs: 1476, groups:  subject:frequency, 268

Fixed effects:
                           Estimate Std. Error      df t value Pr(>|t|)
(Intercept)                 -41.539     10.156 221.843  -4.090 6.03e-05 ***
groupIndian                  58.719     14.072 230.925   4.173 4.26e-05 ***
frequency1.3                 -9.869     14.408 224.514  -0.685  0.49409
frequency1.6                 -6.223     14.601 231.673  -0.426  0.67034
frequency1.9                 -9.079     14.757 227.151  -0.615  0.53900
frequency2.2                -11.489     14.424 225.508  -0.796  0.42659
frequency2.5                  1.250     14.478 226.791   0.086  0.93129
frequency2.8                 11.744     14.487 227.273   0.811  0.41843
frequency3.1                 33.627     14.799 229.688   2.272  0.02400 *
frequency3.4                 40.549     14.881 232.891   2.725  0.00692 **
frequency3.7                 47.784     15.240 251.539   3.135  0.00192 **
groupIndian:frequency1.3    -10.094     20.011 230.019  -0.504  0.61445
groupIndian:frequency1.6    -18.496     20.201 235.007  -0.916  0.36082
groupIndian:frequency1.9    -23.105     20.203 234.137  -1.144  0.25393
groupIndian:frequency2.2    -15.127     19.876 230.377  -0.761  0.44740
groupIndian:frequency2.5    -20.730     20.055 230.108  -1.034  0.30237
groupIndian:frequency2.8    -20.227     20.026 229.716  -1.010  0.31352
groupIndian:frequency3.1    -30.586     20.274 231.743  -1.509  0.13275
groupIndian:frequency3.4    -35.426     20.447 237.672  -1.733  0.08446 .
groupIndian:frequency3.7    -42.973     20.947 248.701  -2.051  0.04127 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Random effects and residuals

What is reported is the variance of mean asynchrony explained by the random intercepts and the residuals

Linear mixed models: Illustration

```
> lmm_mean_async <- lmerTest::lmer(
+   mean_async ~ group * frequency + (1 | subject:frequency),
+   data = df_async_by_task
+ )
> summary(lmm_mean_async)

Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: mean_async ~ group * frequency + (1 | subject:frequency)
   Data: df_async_by_task

REML criterion at convergence: 14731.5

Scaled residuals:
    Min      1Q  Median      3Q     Max
-5.4197 -0.5253  0.0020  0.5799  3.4152

Random effects:
 Groups            Name        Variance Std.Dev.
 subject:frequency (Intercept) 1177.9   34.32
 Residual                       980.2   31.31
Number of obs: 1476, groups:  subject:frequency, 268

Fixed effects:
                          Estimate Std. Error      df t value Pr(>|t|)
(Intercept)                -41.539     10.156 221.843  -4.090 6.03e-05 ***
groupIndian                 58.719     14.072 230.925   4.173 4.26e-05 ***
frequency1.3                -9.869     14.408 224.514  -0.685  0.49409
frequency1.6                -6.223     14.601 231.673  -0.426  0.67034
frequency1.9                -9.079     14.757 227.151  -0.615  0.53900
frequency2.2               -11.489     14.424 225.508  -0.796  0.42659
frequency2.5                 1.250     14.478 226.791   0.086  0.93129
frequency2.8                11.744     14.487 227.273   0.811  0.41843
frequency3.1                33.627     14.799 229.688   2.272  0.02400 *
frequency3.4                40.549     14.881 232.891   2.725  0.00692 **
frequency3.7                47.784     15.240 251.539   3.135  0.00192 **
groupIndian:frequency1.3   -10.094     20.011 230.019  -0.504  0.61445
groupIndian:frequency1.6   -18.496     20.201 235.007  -0.916  0.36082
groupIndian:frequency1.9   -23.105     20.203 234.137  -1.144  0.25393
groupIndian:frequency2.2   -15.127     19.876 230.377  -0.761  0.44740
groupIndian:frequency2.5   -20.730     20.055 230.108  -1.034  0.30237
groupIndian:frequency2.8   -20.227     20.026 229.716  -1.010  0.31352
groupIndian:frequency3.1   -30.586     20.274 231.743  -1.509  0.13275
groupIndian:frequency3.4   -35.426     20.447 237.672  -1.733  0.08446 .
groupIndian:frequency3.7   -42.973     20.947 248.701  -2.051  0.04127 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Random effects and residuals

What is reported is the variance of mean asynchrony explained by the random intercepts and the residuals

Fixed effects, reported as for a simple linear model

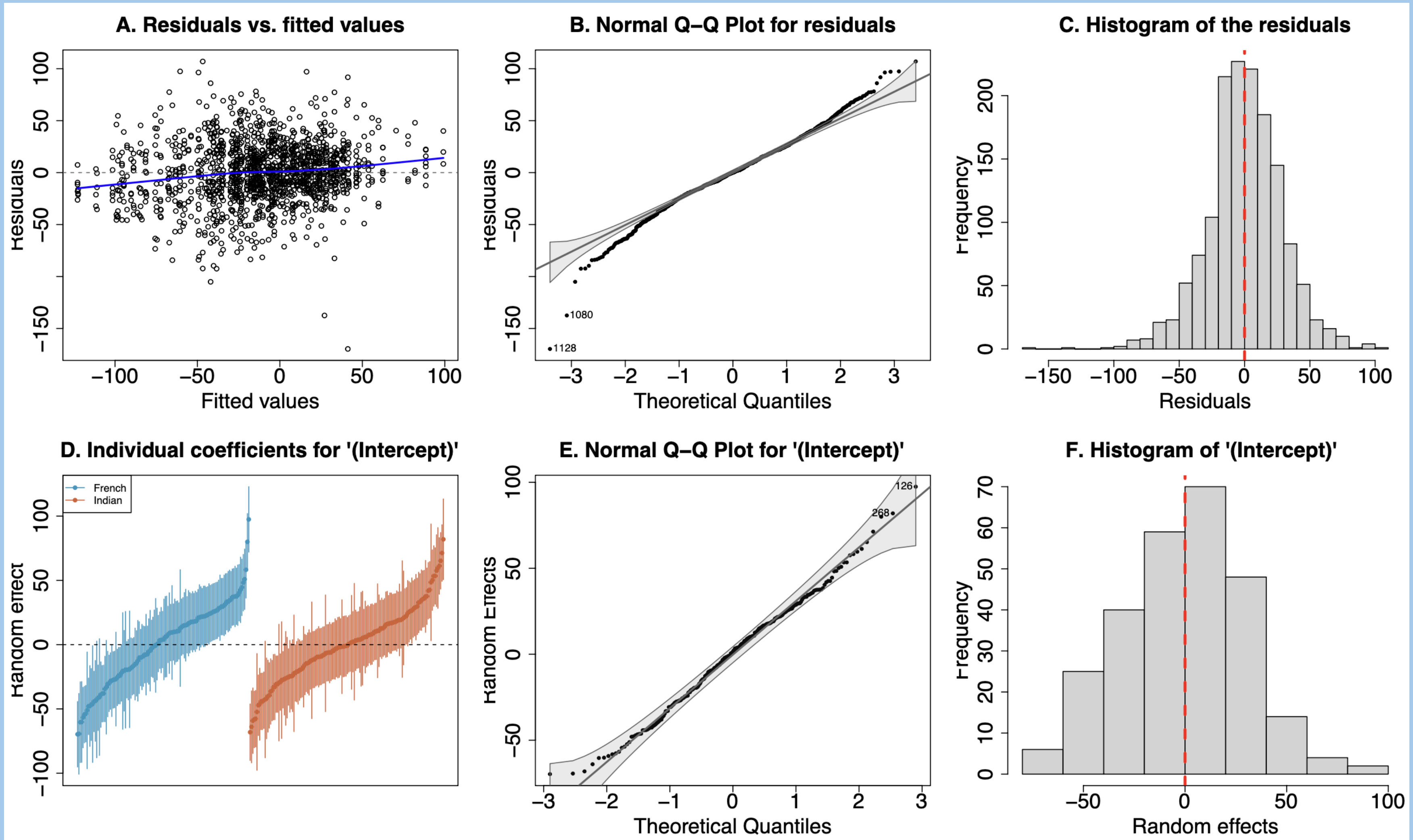
The df are estimated using Satterthwaite approximation (default in lmer)

- (Intercept): value for the French group (baseline) at 1.0 Hz (baseline)
- groupIndian: difference between Indian and baseline group at 1.0 Hz (baseline)
- frequency1.3: difference between 1.3 Hz and baseline for French group (baseline)
- groupIndian:frequency1.3: difference between the Indian group at 1.3 Hz and the baseline

How can I get the estimated marginal means by the model?

French 1.0 Hz = (Intercept)
French 1.3 Hz = (Intercept) + frequency1.3
Indian 1.0 Hz = (Intercept) + groupIndian
Indian 1.3 Hz = (Intercept) + groupIndian + frequency1.3 + groupIndian:frequency1.3

Linear mixed models: Illustration



Linear mixed models: Illustration

Fixed effects

We can analyse the main effects and interaction using F-test

```
> anova(lmm_mean_async, type = 3)

Type III Analysis of Variance Table with Satterthwaite's method

                Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
group            64091   64091     1 236.77 65.3833 3.186e-14 ***
frequency        51811    5757     9 236.15  5.8728 2.160e-07 ***
group:frequency   6299     700     9 236.15  0.7140    0.6959
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Kuznetsova, Brockhoff & Christensen, 2017, *J Stat Soft*

Linear mixed models: Illustration

Fixed effects

We can compute the estimated marginal means for each group at each frequency

```
> emmeans(lmm_mean_async, ~ group | frequency)

frequency = 1.0:
 group   emmean    SE  df lower.CL upper.CL
 French -41.539 10.16 230  -61.550   -21.53
 Indian  17.180  9.74 251   -2.007    36.37

frequency = 1.3:
 group   emmean    SE  df lower.CL upper.CL
 French -51.408 10.22 236  -71.542   -31.27
 Indian  -2.783  9.90 240  -22.281    16.72

...
```

Lenth, 2023

Linear mixed models: Illustration

Fixed effects

We can compute the estimated marginal means for each group at each frequency

```
> emmeans(lmm_mean_async, ~ group | frequency)

frequency = 1.0:
 group   emmean    SE  df lower.CL upper.CL
 French -41.539 10.16 230  -61.550   -21.53
 Indian  17.180  9.74 251   -2.007    36.37

frequency = 1.3:
 group   emmean    SE  df lower.CL upper.CL
 French -51.408 10.22 236  -71.542   -31.27
 Indian  -2.783  9.90 240  -22.281    16.72

...
```

```
> pairs(emmeans(lmm_mean_async, ~ group | frequency))

frequency = 1.0:
 contrast        estimate   SE  df t.ratio p.value
 French - Indian    -58.7 14.1 240  -4.172  <.0001

frequency = 1.3:
 contrast        estimate   SE  df t.ratio p.value
 French - Indian    -48.6 14.2 238  -3.418  0.0007

frequency = 1.6:
 contrast        estimate   SE  df t.ratio p.value
 French - Indian    -40.2 14.5 248  -2.775  0.0059

...
```

Lenth, 2023

Linear mixed models: Illustration

Random effects

Intraclass correlation (ICC): ranges from 0 (random effects conveys no information) to 1

The ICC for the jth random effect (re_j) is:

$$\mathrm{ICC}(re_j) = \frac{\sigma^2_{re_j}}{\sum_{n=1}^{k} \sigma^2_{re_n} + \sigma^2_{\varepsilon}}$$

Gelman & Hill, 2008

We can use the *icc()* function from the *performance* package

```
> icc(lmm_mean_async)

# Intraclass Correlation Coefficient

    Adjusted ICC: 0.546
  Unadjusted ICC: 0.418
```

```
Random effects:
 Groups             Name        Variance Std.Dev.
 subject:frequency (Intercept) 1177.9   34.32
 Residual                        980.2   31.31
```

$$\frac{1177.9}{1177.9 + 980.2} = 0.546$$

Linear mixed models: Illustration

Model fit

Marginal R^2 = proportion of variance explained by the fixed-effects

Conditional R^2 = proportion of variance explained by the fixed-effects and the random-effects

Marginal R^2 ≤ Conditional R^2

```
> performance::r2(lmm_mean_async)

# R2 for Mixed Models

  Conditional R2: 0.652
     Marginal R2: 0.234
```

Main references

Faraway, J. J. (2015). *Linear models with R* (Second edition). CRC Press, Taylor & Francis Group.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

West, B. T., Welch, K. B., & Gałecki, A. T. (2022). *Linear Mixed Models: A Practical Guide Using Statistical Software* (3rd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781003181064

Other references

Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, *33*(2), 261–304. https://doi.org/10.1177/0049124104268644

Long, J. S., & Ervin, L. H. (2000). Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model. *The American Statistician*, *54*(3), 217–224. https://doi.org/10.1080/00031305.2000.10474549

Montgomery, D. C. (2017). *Design and Analysis of Experiments* (Ninth edition). John wiley & sons.

Packages

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Lenth, R. V. (2023) *emmeans: Estimated Marginal Means, aka Least-Squares Means* [R package version 1.9.0]. https://cran.r-project.org/package=emmeans

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*, *6*(60), 3139. https://doi.org/10.21105/joss.03139

R Core Team. (2023). *R: A Language and Environment for Statistical Computing* (Version 4.3.2). R Foundation for Statistical Computing. https://www.R-project.org/

Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Complementary readings

Cox, D. R., & Hinkley, D. V. (1974) *Theoretical Statistics*. Springer US.

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology*, *31*(4), 337–350. https://doi.org/10.1007/s10654-016-0149-3