

# Linear Mixed-Effects Models (aka Statistics III)

Bernd Figner  
b.figner@psych.ru.nl

Week 4: February 29, 2016

1

## Today: *p* values

- **Recap: random slopes**
- **HW/lab/BB-related stuff**
- ***p* values!**
  - Many different roads leading to Rome...
  - Some detours (Type 2 vs 3; testing coefficients vs effects, multiple cores, ...)
- **Homework/lab session**
  - *p* values
  - lmer tutorial

2

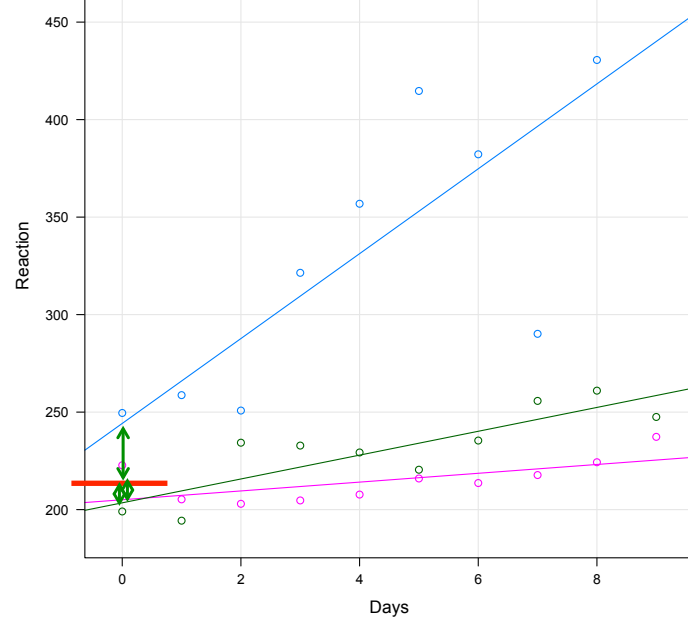
# Introducing our new friend

## Random slopes

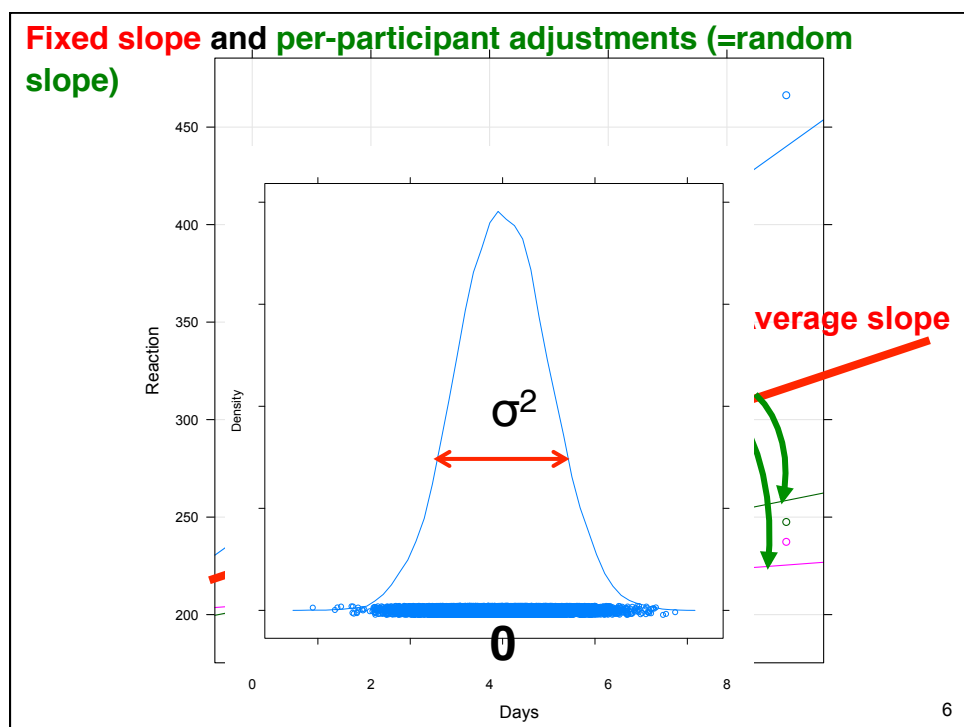
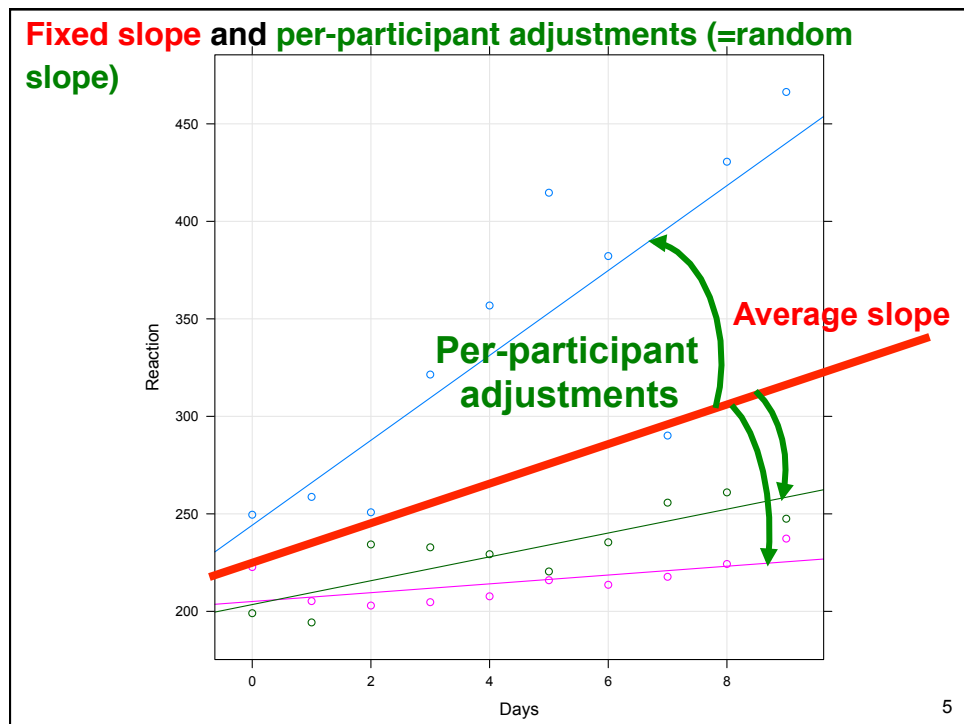
- Example: real data
- sleepstudy

3

**Fixed intercept** and **per-participant adjustments (=random intercept)**



4



## All put together

$$Y_{ij} = (b_0 + u_{0j}) + (b_1 + u_{1j})X_{ij} + \varepsilon_{ij}$$

- **$b_0 + u_{0j}$** 
  - Fixed and random intercept
- **$b_1 + u_{1j}$** 
  - Fixed and random slope
- **$X_{ij}$** 
  - Days predictor
- **$\varepsilon_{ij}$** 
  - Error

**Plus: We learned about "random correlations" (covariances between random effects)**

7

## Today: $p$ values

- **Recap: random slopes**
- **HW/lab/BB-related stuff**
- **$p$  values!**
  - Many different roads leading to Rome...
  - Some detours (Type 2 vs 3; coefficients vs effects, ...)
- **Homework/lab session**
  - $p$  values
  - lmer tutorial

8

## Questions from HW?

- More questions and answers on BlackBoard!
- Warnings when doing profile-based CIs for the cherry pit data
- Was it too little? too much? too easy? too difficult?

9

## Today: $p$ values

- Recap: random slopes
- HW/lab/BB-related stuff
- **$p$  values!**
  - Many different roads leading to Rome...
  - Some detours (Type 2 vs 3; coefficients vs effects, ...)
- Homework/lab session
  - $p$  values
  - lmer tutorial

10

# ***p* values!**

## **What we did so far: Confidence intervals**

- **Fixed effects (i.e., regression coefficients):** How much confidence can we have that the coefficient of interest is different from 0?
  - Factors: Each of the contrasts representing the factor gets their own CI (often not very meaningful)
- **Random effects:** How much confidence can we have that the variance/SD/correlation of interest is different from 0?

### **Thus, for example for fixed effect:**

- Take point estimate (i.e., regression coefficient)
- Quantify the (un)certainty around that estimate
- Can we be sufficiently sure (e.g., 95%) it differs from 0?
- If yes, we say the coefficient is significant (i.e., significantly different from 0, e.g., with  $p < .05$ )

12

## What we do now: $p$ values

### Different idea

- Does the predictor of interest significantly improve model fit?

### Typical approach

- Run model **with** predictor of interest
  - Compute some model fit quantity for that model
- Run model **without** that predictor of interest (otherwise identical)
  - Compute some model fit quantity for that simplified model
- Compare the two model fits
- If the more complex model (the one **with** the predictor) has a significantly better model fit: We conclude that the predictor is significant.

**NOTE: You can compare only nested models**

13

## Sounds simple, right?

- Well, it is

### Except...

- There are a few things that make it a bit complicated sometimes, like:
  - Different methods (similar to different ways to get CIs)
  - Type 2 versus Type 3 tests
  - Tests of coefficients versus tests of effects
    - Factors (with more than 2 levels)
    - Interaction effects including factors
    - Contrast coding scheme matters (e.g., sum-to-zero vs treatment etc)
- But we'll start simple

14

## Sleepstudy hypothesis

- After more nights of sleep deprivation, participants have longer RTs

## Our model from last week

```
s_1 <- lmer(Reaction ~ Days + (1 + Days | f_Subject), data = sleepstudy2)
```

## Hypothesis in terms of model

→ significant fixed effect of Days

- We did it with CIs, now we do it with model comparisons

15

## Model **with** predictor of interest

```
s_1 <- lmer(Reaction ~ Days + (1 + Days | f_Subject), data = sleepstudy2)
```

## Model **without** predictor of interest

```
s_1_noDay <- lmer(Reaction ~ (1 + Days | f_Subject), data = sleepstudy2)
```

## Please note:

- Random slope for Days still in the model!  
→ We want to test significance of fixed effect only  
(*average* effect; not whether there's individual variation)

16



```
summary(s_1_noDay)
[...]
```

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
f_Subject	(Intercept)	651.6	25.53	
	Days	142.2	11.93	-0.18
	Residual	654.9	25.59	

Number of obs: 180, groups: f\_Subject, 18

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	257.76	6.76	38.13

**Looks like no errors/warnings/problems etc, good!**

17

## So how do we compare these two models?

- Actually, there are many many ways...
- Some more recommended, others less
- Not all work in all cases etc
- I will show you several different ways, and we'll have to make some detours along the way
- In the end, I'll give you an overview of the methods with some recommendations
- But I promised we start simple, so let's do that

18

## anova()

You know that from Stats II course

**anova()** does model comparisons

**For lme4 models: Likelihood Ratio Tests (LRTs)**

### NOTES

- LRTs require models to be fit with ML instead of REML  
<http://stats.stackexchange.com/questions/48671/what-is-restricted-maximum-likelihood-and-when-should-it-be-used>
- If you fit your models with REML (as we always do), anova() will re-fit them for you with ML
- The command wants the "smaller" model first, followed by the "larger" model, but it will re-order them for you if you put them in in the wrong order
- You can compare more than 2 models at the same time

19

## anova(s\_1\_noDays, s\_1)

refitting model(s) with ML (instead of REML)

Data: sleepstudy2

Models:

s\_1\_noDays: Reaction ~ (1 + Days | f\_Subject)

s\_1: Reaction ~ Days + (1 + Days | f\_Subject)

	Df	AIC	BIC	logLik	deviance	Chisq	Chi	Df	Pr(>Chisq)
s_1_noDays	5	1785.5	1801.4	-887.74	1775.5				
s_1	6	1763.9	1783.1	-875.97	1751.9	23.537	1	1	1.226e-06***

### Notes

- The test statistic (Chisquare) is difference between the two deviances
- The Dfs for the test: the difference between the Dfs of the two models (in our case the simpler model has 1 Df less)

20

### Write-up example

[...] To determine  $p$  values, we computed Likelihood Ratio Tests using R's `anova` function to compare the model with the respective predictor of interest to the model without the respective predictor of interest (with models being identical otherwise). [...]

Reaction times increased significantly as a function of the number of days with sleep deprivation (Estimate = 10.47(1.55),  $\text{Chisq}(1) = 23.54$ ,  $p < .001$ ).

#### NOTES:

- Feel free to use the Greek letter Chi instead:  $\chi^2$
- You could remove the "Estimate = 10.47(1.55)" part
- You could add the CI for the estimate: "Estimate = 10.47(1.55), 99.9% CI [5.26, 15.39],  $\text{Chisq}(1) = 23.54$ ,  $p < .001$ )"

21

### Bootstrapped Likelihood Ratio Tests

- `PBmodcomp()` from package `pbkrtest`

#### General syntax

`PBmodcomp(LargeModel, SmallModel)`

- Wants 2 models as input: one with the predictor of interest, one without it
- `PBmodcomp` requires models to be fit with ML
- If you fit your models with REML (as we did), `PBmodcomp()` will re-fit them for you with ML
- `PBmodcomp()` wants the "larger" model first, followed by the "smaller" model. **It will NOT reorder them for you! It will do some bootstrapping and even give you some results, but it will not make sense!!!**

22

**Notes continued**

- Default number of simulations = 1000
- If you want a different number:  
PBmodcomp(s\_1, s\_1\_noDays, **nsim = 10000**)
- PBmodcomp runs several tests, including a regular LRT; we're interested in the PBtest  
→ ?PBmodcomp

23

```
s_1Days_PB <- PBmodcomp(s_1, s_1_noDays)
→ Bootstrapping takes a while (40.41 sec in that case)
s_1Days_PB
Parametric bootstrap test; time: 40.41 sec; samples:
1000 extremes: 0;
large : Reaction ~ Days + (1 + Days | f_Subject)
small : Reaction ~ (1 + Days | f_Subject)
      stat df  p.value
LRT    23.516  1 1.239e-06 ***
PBtest 23.516    0.000999 ***
```

24

## Several tests?

`summary(s_1Days_PB)`

Parametric bootstrap test; time: 40.41 sec; samples: 1000 extremes: 0;

large : Reaction ~ Days + (1 + Days | f\_Subject)

small : Reaction ~ (1 + Days | f\_Subject)

	stat	df	ddf	p.value	
PBtest	23.516			0.000999	***
Gamma	23.516			1.488e-05	***
Bartlett	20.779	1.000		5.154e-06	***
F	23.516	1.000	17.185	0.000146	***
LRT	23.516	1.000		1.239e-06	***

If you're curious: ?PBmodcomp

25

## Write-up example

To determine  $p$  values, we used the PBmodcomp function of the package pbkrtest (Halekoh & Højsgaard, 2014)\*, which performs model comparison of nested models using parametric bootstrap methods (i.e., bootstrapped Likelihood Ratio Tests).

Reaction times increased significantly as a function of the number of days with sleep deprivation (Estimate = 10.47(1.55), PBtest = 23.52,  $p < .001$ ).

### NOTES:

- \*Citation information for a package: `citation("pbkrtest")`
- You could remove the "Estimate = 10.47(1.55)" part
- You could add the CI for the estimate: "Estimate = 10.47(1.55), 99.9% CI [5.26, 15.39], PBtest = 23.54,  $p < .001$ )"

26

## Conditional F-tests with Kenward-Roger df correction

- KRmodcomp() from package pbkrtest

### General syntax

```
KRmodcomp(LargeModel, SmallModel)
```

- Only for lmer models, not glmer models
- Wants 2 models as input: one with the predictor of interest, one without it
- KRmodcomp requires models to be fit with REML (sic!)
- If you fit your models with ML, KRmodcomp() will re-fit them for you with REML
- KRmodcomp() wants the "larger" model first, followed by the "smaller" model. **It seems like it reorders them for you, if you enter them in wrong order. (But I wouldn't trust it completely, so better enter them in right order!)**

27

## Our example

```
s_1Days_KR <- KRmodcomp(s_1, s_1_noDays)
s_1Days_KR
F-test with Kenward-Roger approximation; computing time: 0.14 sec.
large : Reaction ~ Days + (1 + Days | f_Subject)
small : Reaction ~ (1 + Days | f_Subject)
      stat    ndf    ddf F.scaling  p.value
Ftest 45.853  1.000 27.997         1 2.359e-07 ***

summary(s_1Days_KR)
F-test with Kenward-Roger approximation; computing time: 0.14 sec.
large : Reaction ~ Days + (1 + Days | f_Subject)
small : Reaction ~ (1 + Days | f_Subject)
      stat    ndf    ddf F.scaling  p.value
Ftest 45.853  1.000 27.997         1 2.359e-07 ***
FtestU 45.853  1.000 27.997         2.359e-07 ***
```

28

## Write-up example

To determine  $p$  values, we used the `KRmodcomp` function of the package `pbkrtest` (Halekoh & Højsgaard, 2014), which performs conditional F tests with degrees of freedom based on Kenward-Roger approximation.

Reaction times increased significantly as a function of the number of days with sleep deprivation (Estimate = 10.47(1.55),  $F(1, 27.997) = 45.85$ ,  $p < .001$ ).

### NOTES:

- You could remove the "Estimate = 10.47(1.55)" part
- You could add the CI for the estimate: "Estimate = 10.47(1.55), 99.9% CI [5.26, 15.39],  $F(1, 27.997) = 45.85$ ,  $p < .001$ )"

29

## So far so good!

### 3 different ways to obtain $p$ values

- LRT using `anova()`
- "Bootstrapped LRT" using `pbkrtest::PBmodcomp()`
- Conditional F test with K-R dfs using `pbkrtest::KRmodcomp()`

### BUT...

- They work only for Type 2 (aka Type II) tests
- Not possible (at least not easy for us) to use them for Type 3 (aka Type III) tests
- Remember the difference between Type 2 and 3?

→ **DETOUR AHEAD!**

30

## Type 2 versus Type 3 Tests

Example: model with main effects A and B and interaction A:B

### Type 2 tests (aka Type II)

For interaction A:B (A:B | A, B)

For main effect A (A | B)

For main effect B (B | A)

### Type 3 tests (aka Type III)

For interaction A:B (A:B | A, B)

For main effect A (A | B, A:B)

For main effect B (B | A, A:B)

### Type 2 vs 3 differs only for models with interactions

<http://goanna.cs.rmit.edu.au/~fscholer/anova.php>  
<https://mcfromnz.wordpress.com/2011/03/02/anova-type-iiiiii-ss-explained>

31

**Many statisticians** prefer **Type 2** tests: If there is a significant interaction, it doesn't make sense to test main effects.

**Psychology (and SPSS default): Type 3** tests are very common; we are interested in main effects even if there are significant interactions.

### In R

- Both are possible, but there is a bug/feature that tries to make it impossible to run models you need for Type 3 tests.
- Luckily, some experts wrote specific functions to make Type 3 tests possible.

**Crucial: For Type 3 tests, you cannot use all contrast types: Sum-to-zero, Helmert, and polynomial are OK; treatment (= dummy) must not be used! For Type 2: All fine**

<http://www.mail-archive.com/r-help@stat.math.ethz.ch/msg69781.html>  
<http://talklab.psy.gla.ac.uk/tvw/catpred/>

32



## R/lmer example code

What lmer models do we need to get  $p$  values?

**(1) Run full model** (main effects and interaction)

```
m_full <- lmer(DV ~ A*B + (1 + A*B | ppID), ...)
```

**(2) Test significance of interaction (same for Type 2 and 3)**

(a) Run reduced model without fixed effect interaction term

```
m_noInteract <- lmer(DV ~ A + B + (1 + A*B | ppID), ...)
```

(b) Compare full model to model without interaction

```
anova(m_noInteract, m_full)
```

```
PBmodcomp(m_full, m_noInteract)
```

```
KRmodcomp(m_full, m_noInteract)
```

33

**(2) Test significance of main effect A**

**Type 2 test**

**We compare a model with A and B to a model with only B**

(a) Model without \*any\* interaction term

```
m_AB <- lmer(DV ~ A + B + (1 + A + B | ppID), ...)
```

(b) Model with only B

```
m_B <- lmer(DV ~ B + (1 + A + B | ppID), ...)
```

(c) Compare the two models

```
anova(m_B, m_AB); or KRmodcomp or PBmodcomp
```

**(3) Test significance of main effect B**

Analogous to test of main effect A

```
anova(m_A, m_AB)
```

34

## How would you do that for a Type 3 tests?

### (2) Test significance of main effect A

We compare full model (A, B, A:B) to a model with B and A:B

(a) Full Model

```
m_full <- lmer(DV ~ A * B + (1 + A * B | ppID), ...)
```

(b) Model from which main effect A has been removed

```
m_noMainA <- lmer(DV ~ B + A:B + (1 + A * B | ppID), ...)
```

(c) Compare the two models

```
anova(m_noMainA, m_full)
```

### BUT

- **R doesn't let you run the model m\_noMainA**
- **Bug/feature:** Interaction term without its respective main effects doesn't make sense (well, if you are a Type 2 person...)
- **Try it out (homework!):** It does something, but not what you want to do

35

## So what should we do? We want Type 3!

### Option 1: There are workarounds

- You can create your own interaction term so that R doesn't notice that the respective main effect is missing; i.e., the interaction term looks like it is a main effect
- You can dig into the R/lme4 machinery, grab the model matrix, and change it manually so R does what you want

36

## Option 2: Thank others that did that for you

Several functions do the complicated stuff for you and run the correct Type 3 tests; **END OF DETOUR!**

- **car::Anova()**
  - F tests with K-R approx.
  - Wald Chiquare tests (not recommended! but fast)
  - Type 2 and Type 3
- **afex:mixed()**
  - F tests with K-R approx.
  - LRT
  - PB
  - Type 3 (also Type 2, but not same as car::Anova())
- others: e.g., package lmerTest
  - We will not cover them in the course

37

## car::Anova() and afex:mixed()

- Both use the pbkrtest package
- Both do some additional nice things for you:
- Both run ALL submodels (i.e., you don't have to create them yourself and then do comparisons)
- Anova() requires that you run your lmer model first and then put the model into the Anova command
- mixed() can replace the lmer command
- mixed() will check whether your continuous predictors are centered; if not, it will tell you so
- mixed() uses automatically sum-to-zero contrasts for all your factors (can be disabled)

38

**How to use Anova()**

```
library(car)
```

```
s_1 <- lmer(Reaction ~ Days + (1 + Days |
f_Subject), data = sleepstudy2)
```

**For K-R F type 3 tests**

```
Anova(s_1, type = 3, test = "F")
```

**Same, but type 2**

```
Anova(s_1, type = 2, test = "F")
```

**Wald Chisquare tests** (not recommended, but fast)

```
Anova(s_1, type = 3, test = "Chisq")
```

39

```
Anova(s_1, type = 3, test = "F")
```

Analysis of Deviance Table (Type III Wald F tests with Kenward-Roger df)

Response: Reaction

	F	Df	Df.res	Pr(>F)
(Intercept)	1357.061	1	25.688	< 2.2e-16 ***
Days	45.853	1	27.997	2.359e-07 ***

```
Anova(s_1, type = 2, test = "F")
```

Analysis of Deviance Table (Type II Wald F tests with Kenward-Roger df)

Response: Reaction

	F	Df	Df.res	Pr(>F)
Days	45.853	1	27.997	2.359e-07 ***

40

### Write-up example

To determine  $p$  values, we computed Type 3 conditional F tests with Kenward-Roger approximation for degrees of freedom as implemented in the `Anova` function of the package `car` (Fox & Weisberg, 2011), which in turn calls the function `KRmodcomp` of the package `pbkrtest` (Halekoh & Højsgaard, 2014).

Reaction times increased significantly as a function of the number of days with sleep deprivation (Estimate = 10.47(1.55),  $F(1, 27.997) = 45.85$ ,  $p < .001$ ).

41

### `mixed()` and the package `afex`

```
install.packages("afex")
library(afex)
```

#### Several useful "convenience" functions

- `mixed()`
- `allFit()`
- plus lots more like `rmANOVAs` plotting, etc!

#### Convenience function?

`mixed()` calls:

- `lme4` to compute mixed models
- and then `pbkrtest` and/or `anova()` to compute  $p$  values
- it can also use `parallel` to use several cores

42

### General syntax

```
m1 <- mixed(DV ~ IV1 + IV2 + (1 + IV2 |
f_group), data = mydata, type = 3, method =
c("KR", "PB", "LRT"), cl = mycluster)
```

### Explanations

- type → Type 3 tests (also does kind of Type 2)
- method → KRmodcomp, PBmodcomp, anova only 1 method at a time!
- cl → cluster created with makeCluster()

### Additional arguments when using method = "PB"

```
args.test = list(nsim = 10000, cl = mycluster)
```

43

### For example our sleep data model: LRT

```
s_1_mixed_LRT <- mixed(Reaction ~ Days + (1 + Days |
f_Subject), data = sleepstudy2, type = 3, method = "LRT")
```

Numerical variables NOT centered on 0 (i.e., interpretation of all main effects might be difficult if in interactions): Days  
 REML argument to lmer() set to FALSE for method = 'PB' or 'LRT'  
 Fitting 2 (g)lmer() models:  
 [..]

### Different types of output (more or less info)

s\_1\_mixed\_LRT → very short, only most essential info  
 s\_1\_mixed\_LRT\$tests → more info, also about the different (sub)models  
 anova(s\_1\_mixed\_LRT) → something in between, often all you need  
 summary(s\_1\_mixed\_LRT) → lmer summary of the full model  
 ?mixed  
 str(s\_1\_mixed\_LRT) → LOTS of stuff in there!!

44

```
anova(s_1_mixed_LRT)
Mixed Model Anova Table (Type 3 tests)

Model: Reaction ~ Days + (1 + Days | f_Subject)
Data: sleepstudy2
Df full model: 6
      Df  Chisq Chi Df Pr(>Chisq)
Days   5 23.537    1 1.226e-06 ***

s_1_mixed_LRT$tests
$Days
Data: sleepstudy2
Models:
fits[[c]]: Reaction ~ 0 + m.matrix[, -2L] + (1 + Days | f_Subject)
full.model: Reaction ~ Days + (1 + Days | f_Subject)
      Df    AIC    BIC logLik deviance Chisq Chi Df Pr(>Chisq)
fits[[c]]   5 1785.5 1801.4 -887.74   1775.5
full.model   6 1763.9 1783.1 -875.97   1751.9 23.537    1 1.226e-06
```

45

## Write-up example

To determine  $p$  values, we computed Type 3 Likelihood Ratio Tests as implemented in the mixed function of the package afex (Singmann, Bolker, & Westfall, 2015).

Reaction times increased significantly as a function of the number of days with sleep deprivation (Estimate = 10.47(1.55),  $\text{Chisq}(1) = 23.54$ ,  $p < .001$ ).

46

**PB**

```
s_1_mixed_PB <- mixed(Reaction ~ Days + (1 + Days |
f_Subject), data = sleepstudy2, type = 3, method = "PB")
```

Numerical variables NOT centered on 0 (i.e., interpretation of all main effects might be difficult if in interactions): Days  
REML argument to lmer() set to FALSE for method = 'PB' or 'LRT'

Fitting 2 (g)lmer() models:

[..]

Obtaining 1 p-values:

[.] → **this takes a while! (bootstrapping)**

```
anova(s_1_mixed_PB)
```

```
Model: Reaction ~ Days + (1 + Days | f_Subject)
```

```
Data: sleepstudy2
```

	Chisq	Chi	Df	Pr(>Chisq)	Pr(>PB)
Days	23.537		1	1.2256e-06	<b>0.000999 ***</b>

47

**Write-up example**

To determine  $p$  values, we computed Type 3 bootstrapped Likelihood Ratio Tests (using 1000 simulations) as implemented in the mixed function of the package afex (Singmann, Bolker, & Westfall, 2015), which in turn calls the function PBmodcomp of the package pbkrtest (Halekoh & Højsgaard, 2014).

Reaction times increased significantly as a function of the number of days with sleep deprivation (Estimate = 10.47(1.55), PBtest = 23.54,  $p < .001$ ).

48



## F test with K-R dfs

```
s_1_mixed_KR <- mixed(Reaction ~ Days + (1 + Days |
f_Subject), data = sleepstudy2, type = 3, method = "KR")
```

Numerical variables NOT centered on 0 (i.e., interpretation of all main effects might be difficult if in interactions): Days

Fitting 2 (g)lmer() models:

[..]

Obtaining 1 p-values:

[.]

```
anova(s_1_mixed_KR)
```

Model: Reaction ~ Days + (1 + Days | f\_Subject)

Data: sleepstudy2

	num	Df	den	Df	F.scaling	F	Pr(>F)
Days	1	27.997			1	45.853	2.359e-07 ***

49

## Write-up example

To determine  $p$  values, we computed Type 3 conditional F tests with Kenward-Roger approximation for degrees of freedom as implemented in the `mixed` function of the package `afex` (Singmann, Bolker, & Westfall, 2015), which in turn calls the function `KRmodcomp` of the package `pbrtest` (Halekoh & Højsgaard, 2014).

Reaction times increased significantly as a function of the number of days with sleep deprivation (Estimate = 10.47(1.55),  $F(1, 27.997) = 45.85$ ,  $p < .001$ ).

50

## Speeding Up

- `mixed()` fits as many models as there are fixed effects (+1)  
→ can take a long time for complex models
- bootstrapping (PB only) also can take a long time
- Easy to use more than 1 CPU to do things in parallel

### **More generally (beyond just `mixed()`):**

- Bootstrapping is S L O W
- Using more than 1 CPU speeds up things A LOT
- How to use several CPUs may differ between Mac/PC and different OS versions → google it for your situation!
- Necessary to install packages (snow, parallel, multicore, ...)
- The required packages change sometimes (different operating systems, different R versions, ...)
- But it's worth the hassle → saves a lot of time!
- **Always leave at least 1 core unused by R!!**

51

## What currently works for me

Mac, OS X 10.9.5

```
install.packages("parallel")
```

```
library(parallel)
```

```
detectCores()
```

→ Tells me that my laptop has 4 cores

52

## Multiple Cores with `mixed()`

### (1) Create the cluster

```
n_cores <- detectCores()
MyCluster <- makeCluster(rep("localhost",
n_cores - 1))
```

### (2) Run your `mixed()` command (LRT, KR, PB)

```
mix1 <- mixed(DV ~ IV + (1 + IV | pp), type =
3, method = "LRT", data = mydata, cl =
MyCluster)
```

→ Each CPU runs a (sub)model! Can save a lot of time.

### (3) Once you're done, stop the cluster

```
stopCluster(MyCluster)
```

53

## Multiple Cores with `mixed("PB")`

- PB first runs all the (sub)models, and then does a lot of bootstrapping to get  $p$  values
- You can use also several CPUs for the bootstrapping

```
mix1 <- mixed(DV ~ IV + (1 + IV | pp), type =
3, method = "PB", data = mydata, cl =
MyCluster, args.test = list(nsim = 1000, cl =
MyCluster))
```

→ **Note:** You can also change the number of simulations (also possible without the cluster command)

### Once you're done, stop the cluster

```
stopCluster(MyCluster)
```

54

## Multiple Cores with PBmodcomp()

```
n_cores <- detectCores()
```

### Create the clusters

```
MyCluster <- makeCluster(rep("localhost",  
n_cores - 1))
```

### Run your model comparison

```
PB_1 <- PBmodcomp(mylarge_m, mysmall_m, nsim =  
1000, cl = MyCluster)
```

### Look at the result

```
PB_1
```

### Stop the cluster

```
stopCluster(MyCluster)
```

55

## Multiple Cores with:

### confint.merMod

```
confint(s_1, method = "boot", parallel =  
"multicore", ncpus = 3)
```

```
confint(s_1, method = "boot", parallel =  
"snow", ncpus = 3)
```

### bootMer()

```
boot_s_1 <- bootMer(s_1, FUN_bootMer,  
parallel = "multicore", ncpus = 3)
```

```
boot_s_1 <- bootMer(s_1, FUN_bootMer,  
parallel = "snow", ncpus = 3)
```

56

## An Overview of what we covered

### (1) `confint()` → CIs → *p* values

- ~~Wald~~ (untrustworthy)
  - Likelihood profile (method = "profile")
  - bootstrapping (method = "boot"); `bootMer()`
- Good, but only for coefficients, not effects

### (2) Conditional F-tests with K-R dfs

- `pbkrtest::KRmodcomp()`; Type 2
  - `car::Anova(test="F")`; Type 2 and 3
  - `afex::mixed(method = "KR")`; Type 3 (and 2)
- Often good (not available for generalized mixed models; might be available also for glmer in the near future?)

57

### (3) Bootstrapped Likelihood Ratio Tests

- `pbkrtest::PBmodcomp()`; Type 2
  - `afex::mixed(method = "PB")`; Type 3 (and 2)
- Good

### (4) Likelihood Ratio Tests

- `anova()`; Type 2
  - `afex::mixed(method = "LRT")`; Type 3 (2)
- Some debate whether good or not

### (5) Wald Chisquare Test

- `car::Anova(test="Chisq")`; Type 2 and 3
- Bad

58

## Significance Tests

- **Typically only for fixed effects of interest**
- **Several options** (we encountered 5)
- Tests of **coefficients** vs. test of **effects**
  - **Coefficients** → like regression: coefficient significantly different from 0?
  - **Effects** → For factors with 3 or more levels and interactions involving factors  
→ like ANOVA: Is the **whole** factor significant?  
Is the **whole** interaction term significant?
- **Post-hoc tests** (e.g., multiple pairwise comparisons)  
→ later in the course

59

**Coefficients:** familiar from regression framework

### Is the coefficient significantly different from 0?

- Continuous predictors, e.g., Days (1 coeff)
- Interactions between 2 continuous predictors (1 coeff)
- Factors with 2 levels (1 coeff)
- Interactions between continuous predictor and 2-level factor (1 coeff)
- Factors with 3 or more levels → Remember?
  - Factors with 3 or more levels (e.g., k levels)...
  - ... are represented as k-1 contrasts
  - Each contrast can be tested for significance, e.g.:
    - Level 1 vs. overall average
    - Level 2 vs. overall average
  - Different coding schemes (sum-to-zero, treatment, Helmert, ...): different contrasts, different interpretation

60

**„Effects:“** familiar from ANOVA framework

**Example: A factor with 3 (or more) levels, e.g.,  
Mood induction: happy, neutral, sad**

- Is the **whole factor** significant?  
→ Does the mood induction influence the DV?
- Interactions between categorical factors  
→ Is the **whole interaction** significant?  
Is the effect of mood induction different for females vs. males?

BTW: After identifying a significant effect of a "whole" factor or "whole" interaction, we want to further investigate where this comes from (e.g., with post-hocs). This is again something different and will be discussed later in the course...

61

## Why is this distinction relevant?

- **Confidence intervals:** Only tests of coefficients possible
- **Model comparisons:** Test of regression coefficients and effects possible

62

## Last $p$ value slide for today

→ **Many** different methods (usually similar results)

(Douglas Bates: <https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html>)

- More available than we are going to cover
- We focus on methods recommended as being reliable

## List ordered according to recommendations

<http://glmm.wikidot.com/faq>

**newer version (will move to here in the future):**

<https://rawgit.com/bbolker/mixedmodels-misc/master/glmmFAQ.html>

63

## Today: $p$ values

- Recap: random slopes
- HW/lab/BB-related stuff
- $p$  values!
  - Many different roads leading to Rome...
  - Some detours (Type 2 vs 3; coefficients vs effects, ...)
- **Homework/lab session**
  - $p$  values
  - lmer tutorial

64



## Today's Homework/Lab Session

### Instructions: BlackBoard → Week 4

Practice computing  $p$  values for a mixed model, using all the methods that we discussed today. Plus some reading/R'ing, also in preparation for next week.

### (1) The Cherry Pit Spitting Data

Compute  $p$  values for the model with gender, trial, and their interaction

- Type II test of the interaction with anova, KRmodcomp, PBmodcomp
- Try (and fail!) doing a type III main effect test with anova
- Type III tests of main effects and interaction with Anova and mixed
- Practice your write-up skills

65

### (2) The lmer tutorial

- Part 2 of the tutorial by Bodo Winter

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499. [<http://arxiv.org/pdf/1308.5499.pdf>]

- available here:

[http://www.bodowinter.com/tutorial/bw\\_LME\\_tutorial2.pdf](http://www.bodowinter.com/tutorial/bw_LME_tutorial2.pdf)

- Data set available here:

[www.bodowinter.com/tutorial/politeness\\_data.csv](http://www.bodowinter.com/tutorial/politeness_data.csv)

- Do all the stuff yourself in R (the same way you did this for part 1 on linear regression)

66

### **What do you have to hand in?**

- Answers to questions, output, figures, R code
- Paste everything into a word doc, print it and
- bring it to the lab session on March 7

67

### **Questions or comments?**

**That's it for today's lecture!**

**See you in the basement at 13:45**

**SP A -1.55A+B**