

Linear Mixed-Effects Models

(aka Statistics III)

Bernd Figner

b.figner@psych.ru.nl

Week 3: February 22, 2016

1

Today: More Theory and Models

- HW/recap/follow-up last week
- Random slopes
 - An example data set
 - Some theory
 - A model with random intercept and random slope
 - More plots
- Homework/lab session
 - My first interaction!

2

Questions from HW?

- Was it too little? too much? too easy? too difficult?
- Very few questions on BlackBoard
- Everything clear and simple or too complicated?

3

Cherry Pit Spitting

- How to add gender to the model?
 - Random intercept? Fixed effect? Random slope?
- Why not random intercept, gender like a grouping variable?
 - Data are still grouped per participant; thus: participant as random intercept

A first rule of thumb (more about this in coming weeks)

If you want to test the significance of a predictor, you probably want to model it as **fixed effect** (sometimes fixed effect **and** random slope), but not as a random intercept

4

Second level model

(participants)

- b_{0j}
 - Intercept for participant j
 - More general: As in linear regression, the intercept represents the value when all predictors are 0
- b_0
 - Overall intercept
 - typically grand mean (across all trials and participants)
- u_j
 - error term for participant j
 - deviation of participant j's mean from the grand mean
 - assumed to be: $N(0, \sigma^2)$

Only true if you use sum-to-zero contrasts

More general: As in linear regression, the intercept represents the value when all predictors are 0

5

Sum-to-zero contrasts

```
options(contrasts = c("contr.sum",
"contr.poly"))
```

- This statement sets the contrasts for ALL your factors (until you close R...)
- First part in the expression (contr.sum): sum-to-zero contrasts for all your **unordered** factors
- Second part (contr.poly): polynomial contrasts for all your **ordered** factors
- I have this statement at the top of all my R scripts, to make sure I run it before I run any models

6

lmer: General Syntax

```
MyModel <- lmer(DV ~ IV1 + IV2 + (1 +  
IV1 | GroupingFactor), data = MyData)
```

- The stuff between parentheses: **Random effects**
- Random effects handle the non-independence in the data
- The **GroupingFactor** variable after the **|** is the "clustering" aka "grouping" variable (e.g., participants; or classes in the case of "students nested in classes")
- There can be more than one grouping factor, but we start simple, with only one...

7

Lab Computers

- At least some computers couldn't install package **pbkrtest**, and **car** didn't work either (because it requires **pbkrtest**)
- It seems the problem was that R version 3.2.2 was installed (and not the most recent, i.e., version 3.2.3)
- IT people said they now installed the newest R version
- Thus, please check R version (RStudio version)
- If there are problems, please let me know!

8

Today: More Theory and Models

- HW/recap/follow-up last week
- **Random slopes**
 - An example data set
 - Some theory
 - A model with random intercept and random slope
 - More plots
- Homework/lab session
 - My first interaction!

9

Introducing our new friend

Random slopes

→ Example: real data

10

Sleep Study

```
library(lme4)
```

```
?sleepstudy
```

Sleep Deprivation Study ($N = 18$)

- The average reaction time on each of 10 days of subjects in a sleep deprivation study
- On day 0 the subjects had their normal amount of sleep
- Starting that night they were restricted to 3 hours of sleep per night; this was done for 9 consecutive nights
- The observations represent the average reaction time on a series of tests given each day to each subject

11

Hypothesis

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

Our analysis?

- If it were linear regression:
 $\text{lm}(\text{Reaction} \sim \text{Days}, \text{data} = \text{sleepstudy})$
- But we have non-independent data!
- Each participant contributes 10 data points
 → mixed models!

12

Let's have a look!

sleepstudy[1:20,]

	Reaction	Days	Subject
1	249.5600	0	308
2	258.7047	1	308
3	250.8006	2	308
4	321.4398	3	308
5	356.8519	4	308
6	414.6901	5	308
7	382.2038	6	308
8	290.1486	7	308
9	430.5853	8	308
10	466.3535	9	308
11	222.7339	0	309
12	205.2658	1	309
13	202.9778	2	309
14	204.7070	3	309
15	207.7161	4	309
16	215.9618	5	309
17	213.6303	6	309
18	217.7272	7	309
19	224.2957	8	309
20	237.3142	9	309

13

What's the data structure?

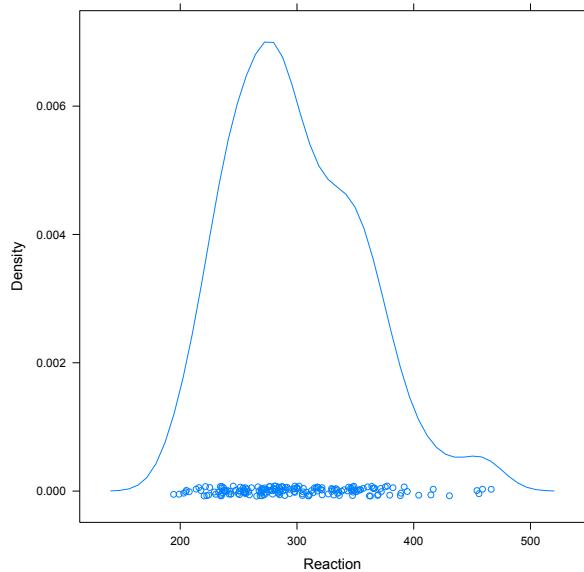
with(sleepstudy, table(Subject, Days))

Subject	Days									
	0	1	2	3	4	5	6	7	8	9
308	1	1	1	1	1	1	1	1	1	1
309	1	1	1	1	1	1	1	1	1	1
310	1	1	1	1	1	1	1	1	1	1
330	1	1	1	1	1	1	1	1	1	1
331	1	1	1	1	1	1	1	1	1	1
332	1	1	1	1	1	1	1	1	1	1
333	1	1	1	1	1	1	1	1	1	1
334	1	1	1	1	1	1	1	1	1	1
335	1	1	1	1	1	1	1	1	1	1
337	1	1	1	1	1	1	1	1	1	1
349	1	1	1	1	1	1	1	1	1	1
350	1	1	1	1	1	1	1	1	1	1
351	1	1	1	1	1	1	1	1	1	1
352	1	1	1	1	1	1	1	1	1	1
369	1	1	1	1	1	1	1	1	1	1
370	1	1	1	1	1	1	1	1	1	1
371	1	1	1	1	1	1	1	1	1	1
372	1	1	1	1	1	1	1	1	1	1

14

Distribution of the DV

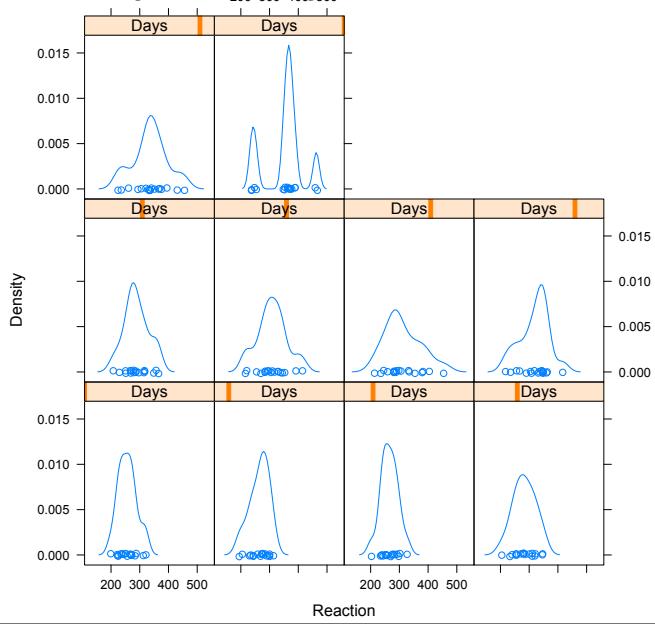
```
with(sleepstudy, densityplot(Reaction))
```



15

Effect of Days on Reaction?

```
with(sleepstudy, densityplot(~ Reaction | Days))
```

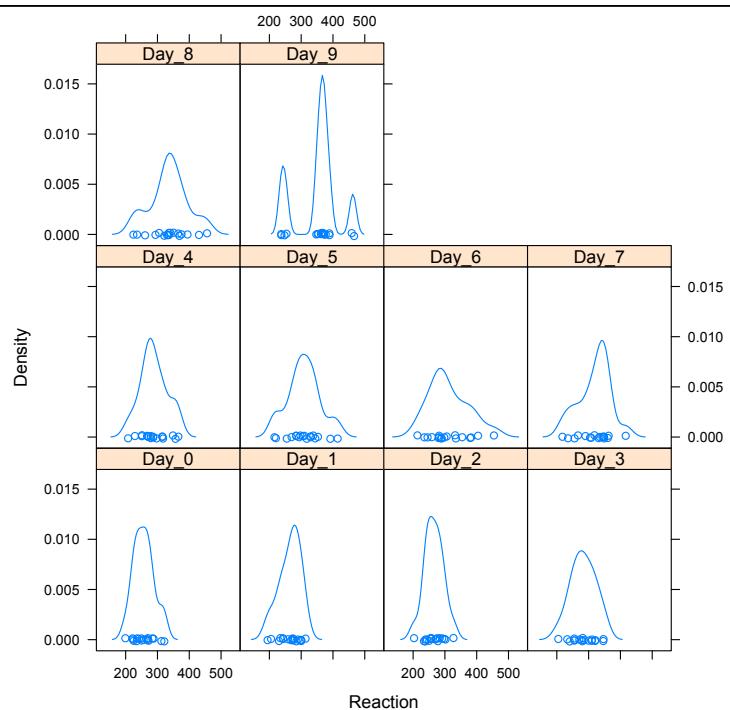


16

Turn Days into factor, just for plotting

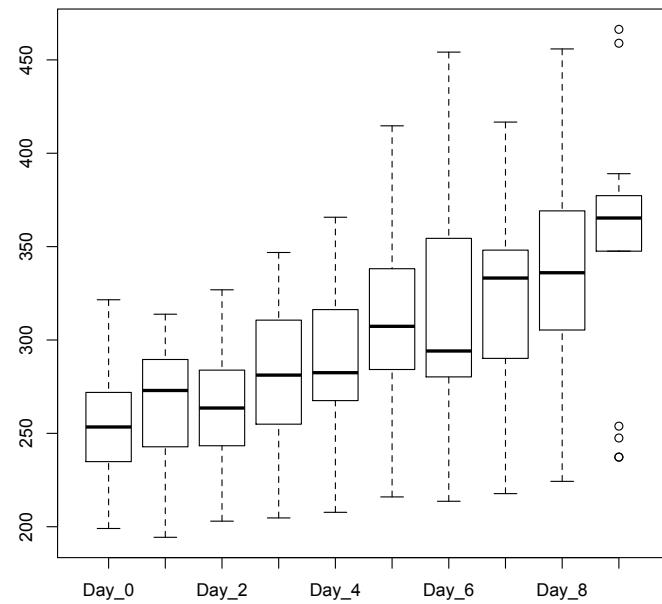
```
sleepstudy2 <- sleepstudy  
  
sleepstudy2$f_Days <-  
as.factor(paste("Day", sleepstudy2$Days,  
sep = '_'))  
  
with(sleepstudy2, densityplot(~ Reaction |  
f_Days))
```

17



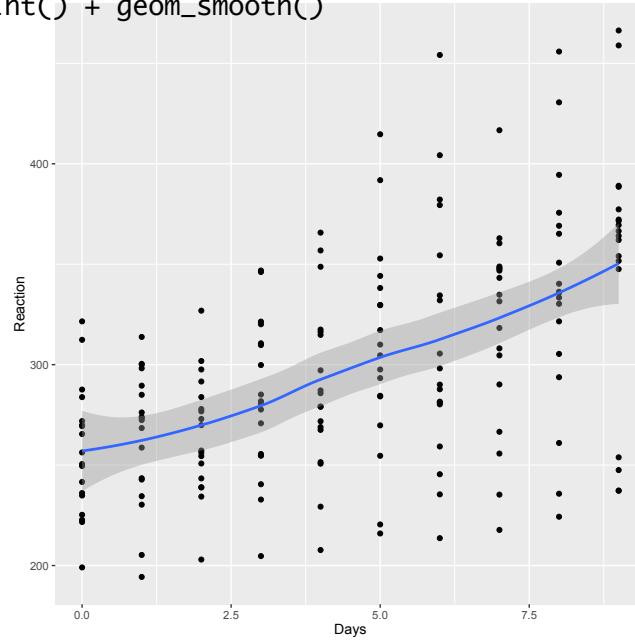
18

```
with(sleepstudy2, boxplot(Reaction ~ f_Days))
```



19

```
ggplot(data = sleepstudy2, aes(x = Days, y = Reaction)) +  
  geom_point() + geom_smooth()
```



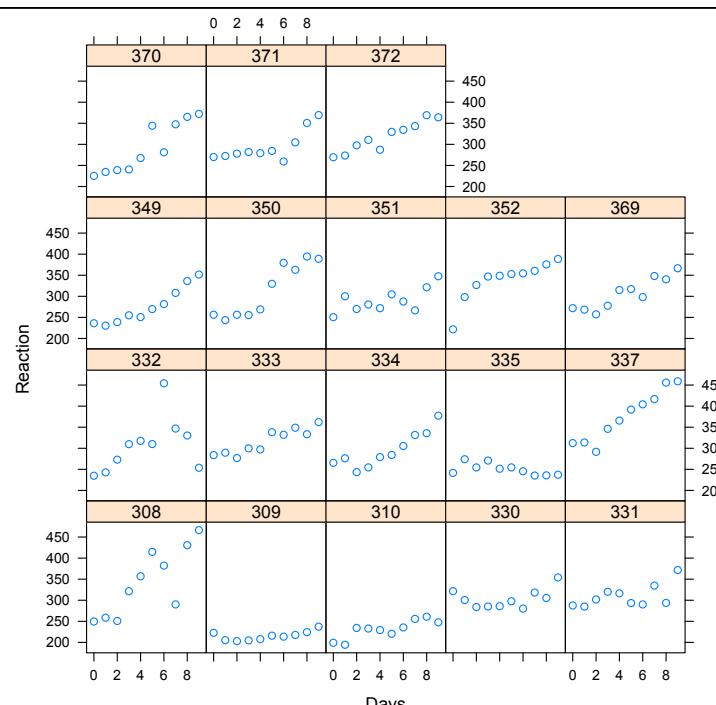
20

Lots of variation -- do participants differ or show similar patterns?

Another new friend: the xyplot()

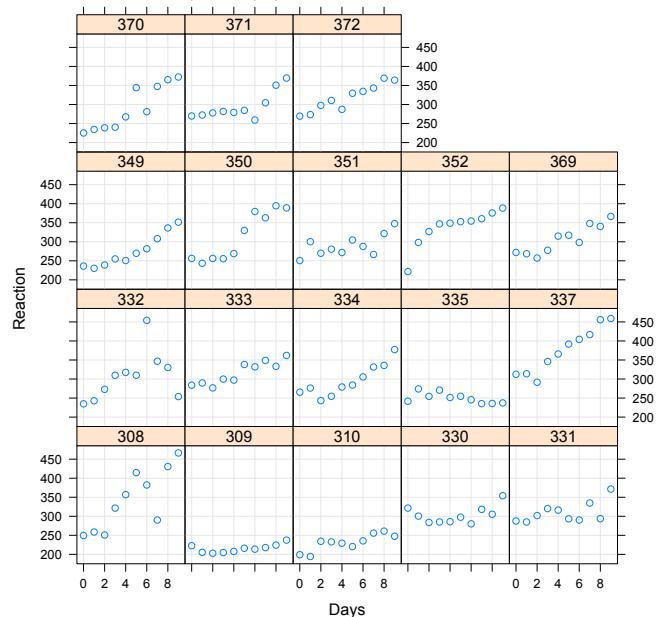
```
xyplot(Reaction ~ Days | Subject, data  
= sleepstudy2)
```

21



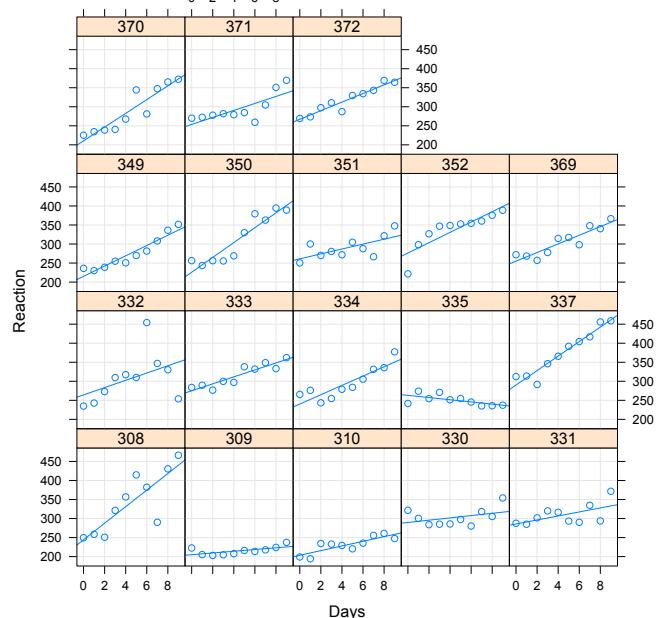
22

```
xyplot(Reaction ~ Days | Subject, data =
sleepstudy2, type = c('g', 'p'))
```



23

```
xyplot(Reaction ~ Days | Subject, data =
sleepstudy2, type = c('g', 'p', 'r'))
```



24

Previous slide

→ Panels ordered according to participant code
(often meaningless)

Order them differently:

By intercept

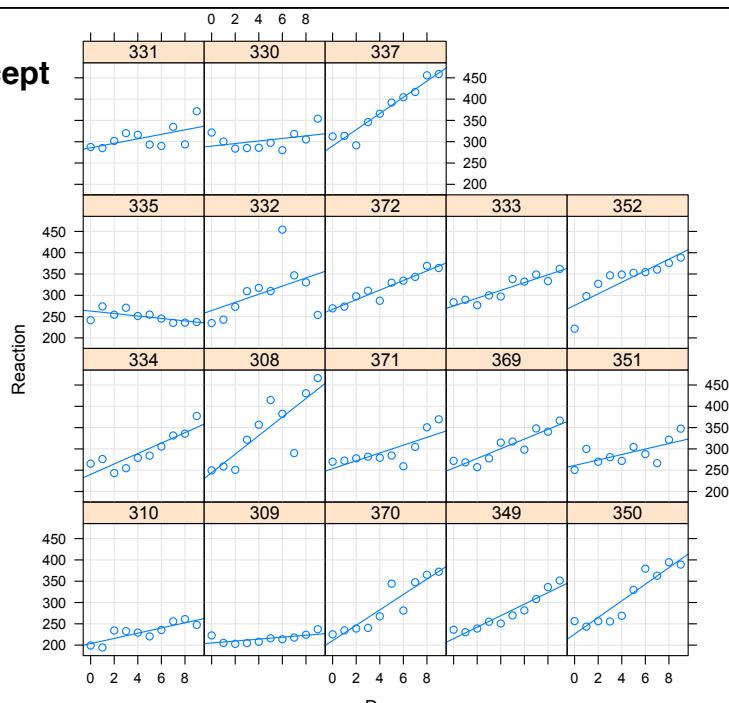
```
xyplot(Reaction ~ Days | Subject, data =
  sleepstudy2, type = c('g', 'p', 'r'),
  index.cond = function(x,y) coef(lm(y ~ x))[1])
```

By slope

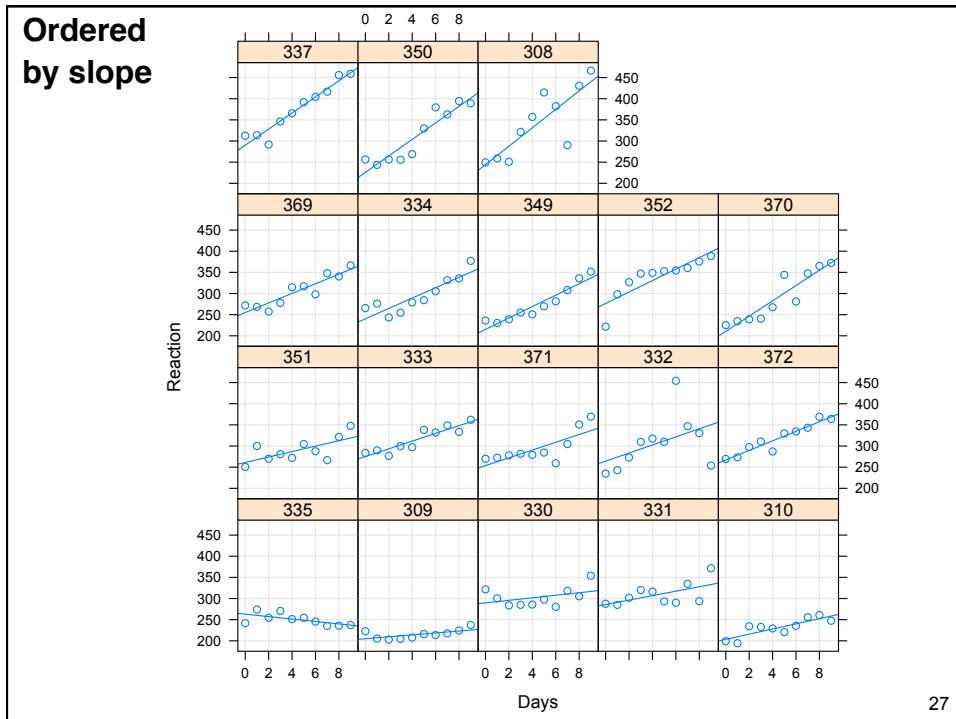
```
xyplot(Reaction ~ Days | Subject, data =
  sleepstudy2, type = c('g', 'p', 'r'),
  index.cond = function(x,y) coef(lm(y ~ x))[2])
```

25

Ordered by intercept



26



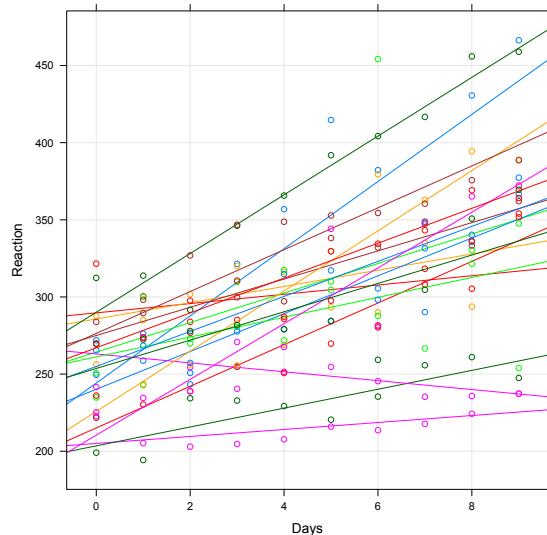
Why am I telling you all this?

Relevant for our models!

- Participants differ in their intercepts
→ **Random intercept**
 - **Participants differ in their slopes**
 - The effect of IV (Days) on DV (Reaction) differs from participant to participant
- **Random slope**

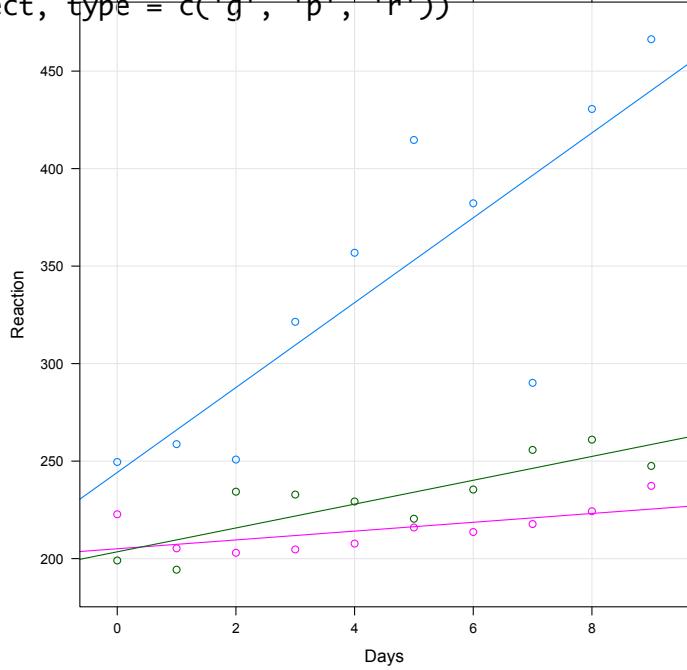
Plotting all participants data into 1 figure

```
xyplot(Reaction ~ Days, data = sleepstudy2, group =
  Subject, type = c('g', 'p', 'r'))
```



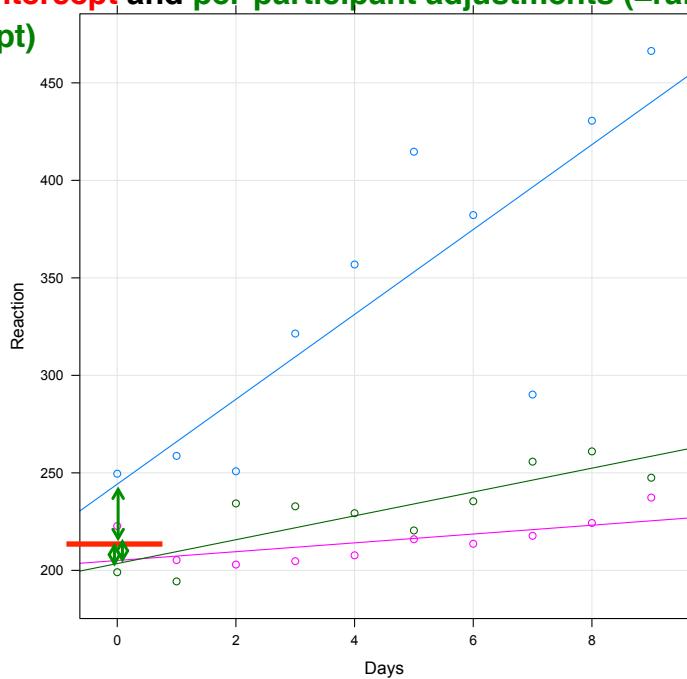
29

```
xyplot(Reaction ~ Days, data = sleepstudy2[1:30,], group =
  Subject, type = c('g', 'p', 'r'))
```



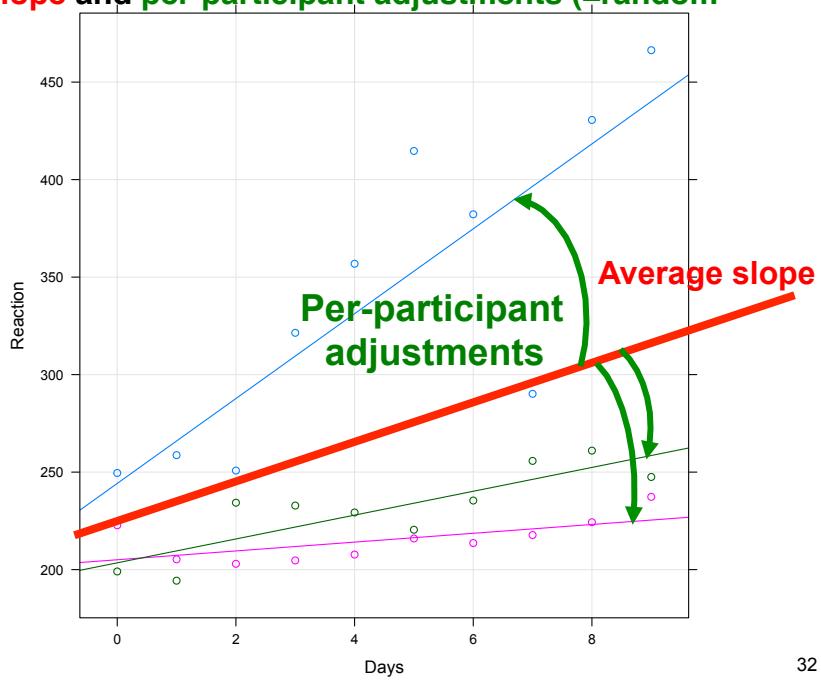
30

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

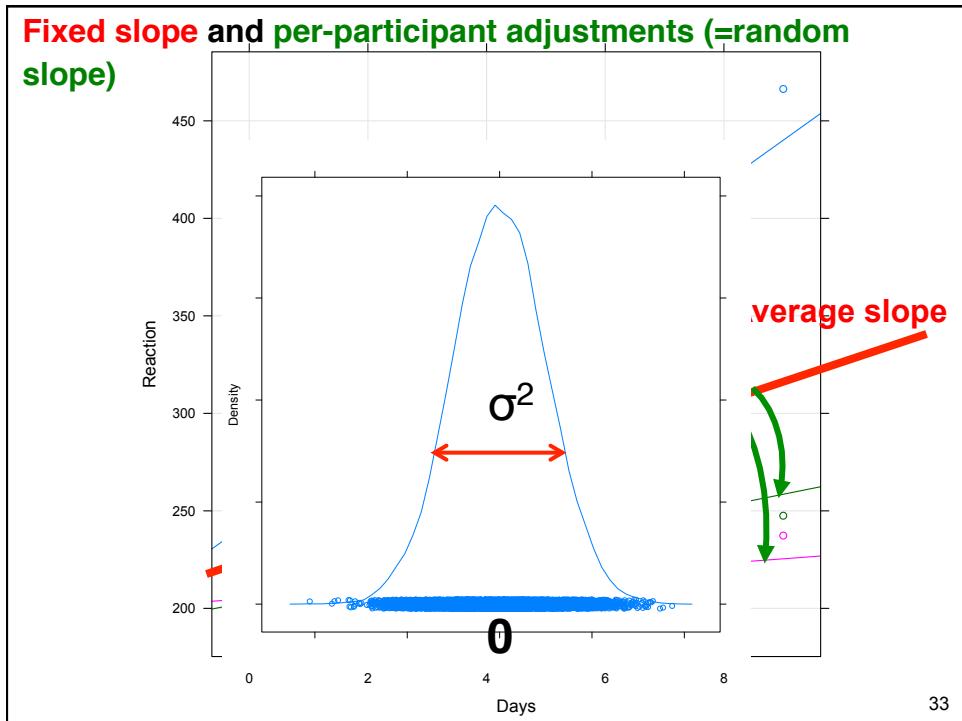


31

Fixed slope and per-participant adjustments (=random slope)



32



Some formulas

34

If it were simple linear regression

$$Y_i = b_0 + b_1 X_i + \varepsilon_i$$

- Y_i
– Reaction time of participant i
- b_0
– Intercept (average for Day 0)
- b_1
– Days coefficient (X: 0, 1, 2, 3, 4, ..., 9)
- ε_i
– Error

35

Mixed-model extension

$$Y_{ij} = b_{0j} + b_{1j} X_{ij} + \varepsilon_{ij}$$

- b_{0j}
– Intercept of pp j
- b_{1j}
– Days coefficient of participant j
- X_{ij}
– Value of the Day predictor for day i of participant j
- ε_{ij}
– Error

36

Second level formulas (participant-level):

$$b_{0j} = b_0 + u_{0j}$$

- b_{0j}
– Intercept of participant j
- b_0
– Overall intercept
- u_{0j}
 - Deviation of participant j's intercept from the overall intercept
 - "Error term" for participants' intercepts
 - Assumed to be: $N(0, \sigma^2)$

37

Second level formulas (participant-level):

$$b_{1j} = b_1 + u_{1j}$$

- b_{1j}
– Slope of participant j
- b_1
– Overall slope → **fixed slope**
- u_{1j}
 - Deviation of participant j's slope from the overall slope
→ **random slope**
 - "Error term" for participants' slopes
 - Assumed to be: $N(0, \sigma^2)$

38

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
- ε_{ij}
– Error

39

Let's run this model

Hypothesis

After more nights of sleep deprivation,
participants have longer RTs

40

Typical Analysis Steps

- Write down hypotheses and/or research questions
- Write down model syntax; express hypotheses/research questions in terms of significant and non-significant effects and interactions
- Collect your data
- Understand your data: plots and/or descriptive statistics
- Run your model
- Check the summary output and do model diagnostics
- Compute *p* values and/or confidence intervals
- Report your model and results in text and figures

41

General lmer syntax

```
MyModel <- lmer(DV ~ IV1 + IV2 +
(1 + IV1|GroupingFactor), data = MyData)
```

IV1 fixed slope and random slope

IV2 fixed slope only

Thus for our analysis (we have only 1 IV)

DV = Reaction

IV1 = Days

GroupingFactor = Subject

42

But first: Turn Subject into explicit factor

```
sleepstudy2$f_Subject <- as.factor(paste("pp",
sleepstudy2$Subject, sep = "_"))
```

Should we center (or standardize) the continuous predictor Days?

Our model

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

Hypothesis in terms of model

→ significant fixed effect of Days

43

Typical Steps

- Write down hypotheses and/or research questions
- Write down model syntax; express hypotheses in terms of significant and non-significant effects and interactions
- Collect your data
- Understand your data: plots and/or descriptive statistics
- **Run your model**
- **Check the summary output** and do model diagnostics
- Compute p values and/or confidence intervals
- Report your model and results in text and figures

44

```

> summary(s_1)
Linear mixed model fit by REML ['lmerMod']
Formula: Reaction ~ Days + (1 + Days | f_Subject)
Data: sleepstudy2

REML criterion at convergence: 1743.6

Scaled residuals:
    Min     1Q Median     3Q    Max 
-3.9536 -0.4634  0.0231  0.4634  5.1793 

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 f_Subject (Intercept) 612.09   24.740  
          Days         35.07   5.922   0.07 
 Residual      654.94   25.592  
Number of obs: 180, groups: f_Subject, 18

Fixed effects:
            Estimate Std. Error t value
(Intercept) 251.405    6.825  36.84 
Days         10.467    1.546   6.77 

Correlation of Fixed Effects:
  (Intr) 
Days -0.138

```

45

```

Linear mixed model fit by REML ['lmerMod']
Formula: Reaction ~ Days + (1 + Days | f_Subject)
Data: sleepstudy2

REML criterion at convergence: 1743.6

Scaled residuals:
    Min     1Q Median     3Q    Max 
-3.9536 -0.4634  0.0231  0.4634  5.1793 


```

46

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
f_Subject	(Intercept)	612.09	24.740	
	Days	35.07	5.922	0.07
Residual		654.94	25.592	

Number of obs: 180, groups: f_Subject, 18

47

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
f_Subject	(Intercept)	612.09	24.740	
	Days	35.07	5.922	0.07
Residual		654.94	25.592	

Number of obs: 180, groups: f_Subject, 18

How many random effects do we have in our model?

- Random intercept (varying over participants)
- Random slope for day (varying over participants)
- Random correlation: Covariance between random intercept and random slope

48

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	251.405	6.825	36.84
Days	10.467	1.546	6.77

Correlation of Fixed Effects:

(Intr)
Days -0.138

What does "Correlation of Fixed Effects" mean?

Good question...

<http://stats.stackexchange.com/questions/57240/how-do-i-interpret-the-correlations-of-fixed-effects-in-my-glmer-output>

49

Correlation of Fixed Effects:

(Intr)
Days -0.138

Douglas Bates (main lme4 developer)

It is an approximate correlation of the estimator of the fixed effects. (I include the word "approximate" because I should but in this case the approximation is very good.) I'm not sure how to explain it better than that. Suppose that you took an MCMC sample from the parameters in the model, then you would expect the sample of the fixed-effects parameters to display a correlation structure like this matrix.

How to suppress in output: `summary(mymodel, cor = FALSE)`

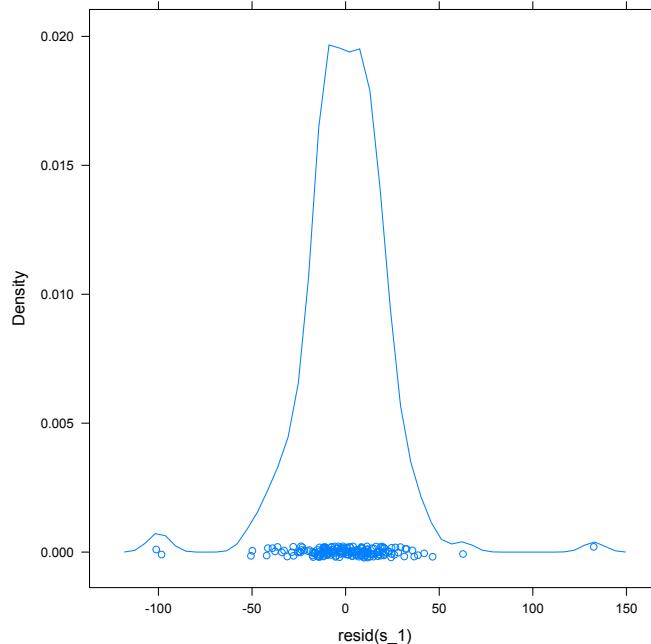
50

Typical Steps

- Write down hypotheses and/or research questions
- Write down model syntax; express hypotheses in terms of significant and non-significant effects and interactions
- Collect your data
- Understand your data: plots and/or descriptive statistics
- Run your model
- Check the summary output and do **model diagnostics**
- Compute p values and/or confidence intervals
- Report your model and results in text and figures

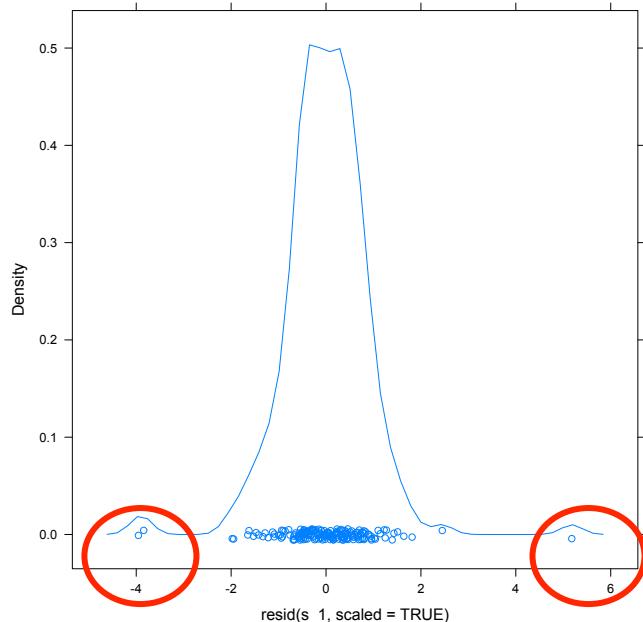
51

`densityplot(resid(s_1))`



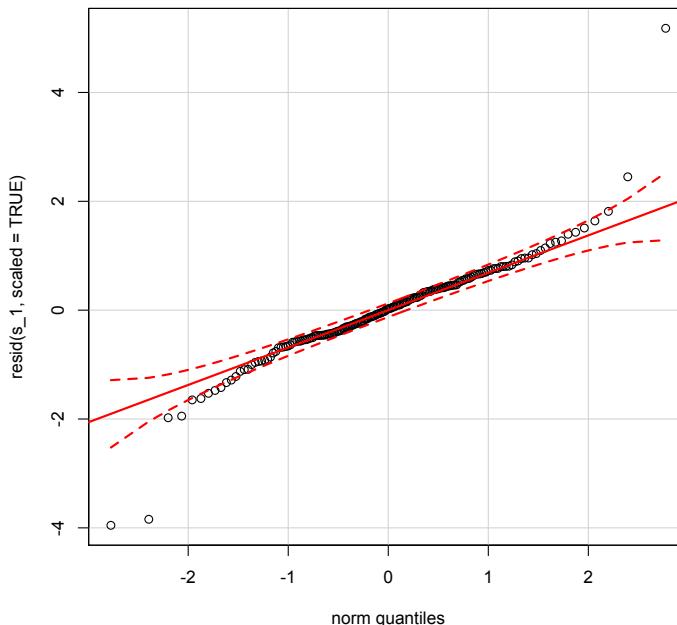
52

```
densityplot(resid(s_1, scaled = TRUE))
```

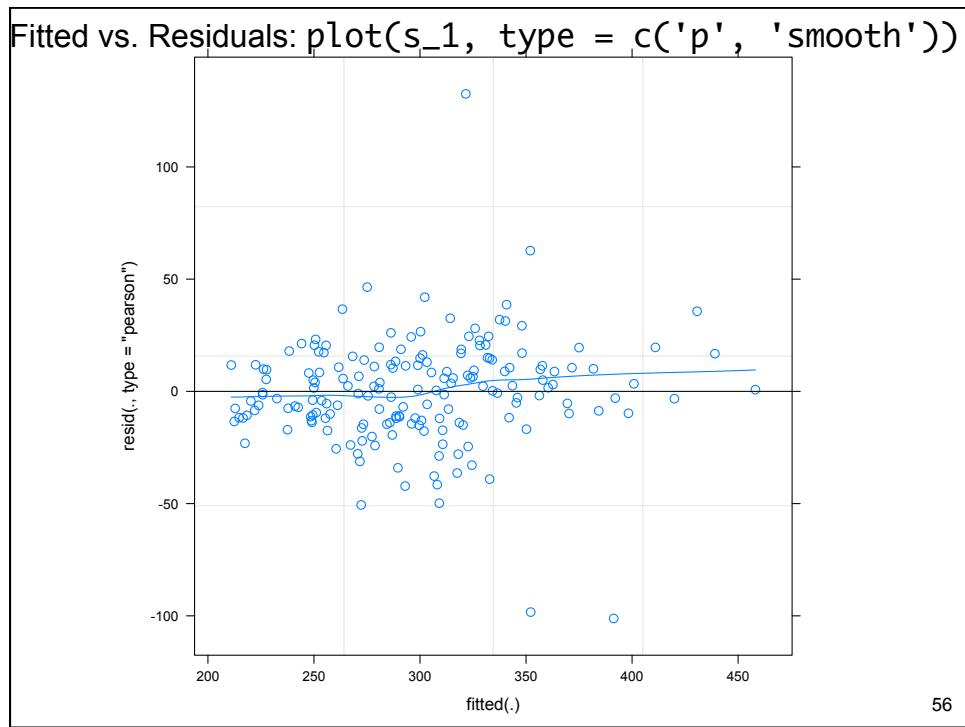
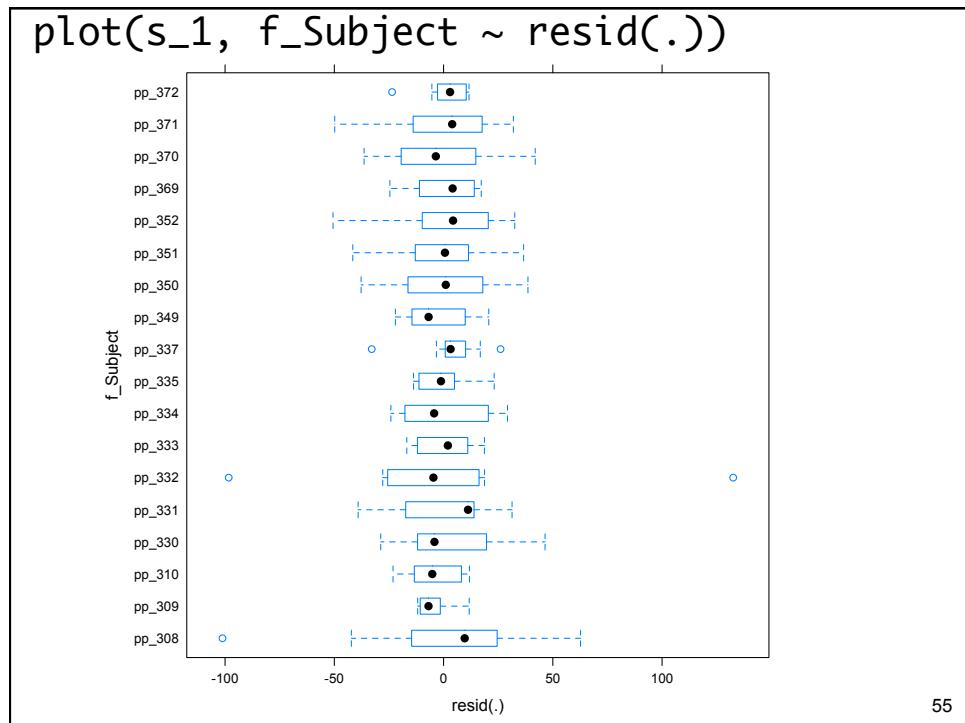


53

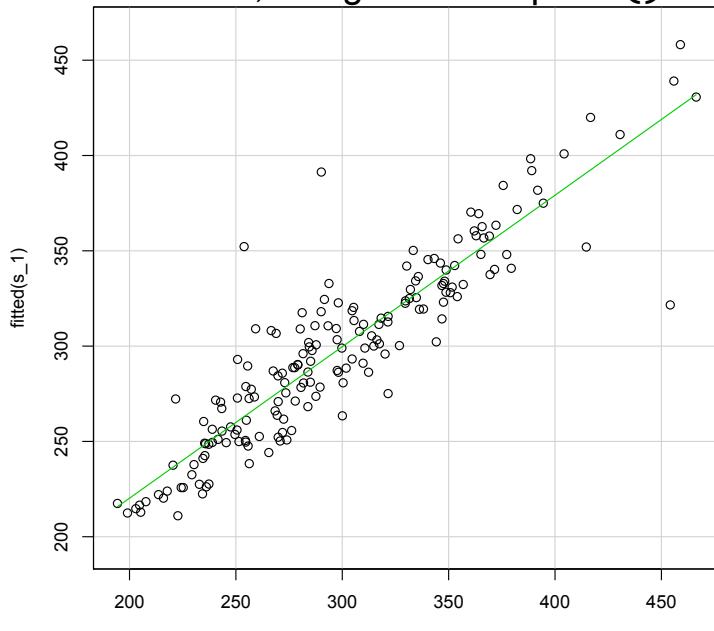
```
from car: qqPlot(resid(s_1, scaled = TRUE))
```



54



Observed vs. fitted, using scatterplot() from car



57

Typical Steps

- Write down hypotheses and/or research questions
- Write down model syntax; express hypotheses/research questions in terms of significant and non-significant effects and interactions
- Collect your data
- Understand your data: plots and/or descriptive statistics
- Run your model
- Check the summary output and do model diagnostics
- **Compute p values and/or confidence intervals**
- Report your model and results in text and figures

58

Hypothesis

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

Profile-based CIs

	0.05 %	99.95 %
[...]		
Days	4.6783334	16.25624

Conclusion

Days effect: $p < .001$, 99.9% CI [4.68, 16.25]

59

Bootstrapping (10,000 simulations): Days $p < .001$

```
s_1_boot <- bootMer(s_1, FUN_bootMer, nsim = 10000, type = "parametric")
Warning messages:
1: In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
  unable to evaluate scaled gradient
2: In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
  Model failed to converge: degenerate Hessian with 1 negative eigenvalues
Warning messages:
1: In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
  unable to evaluate scaled gradient
2: In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
  Model failed to converge: degenerate Hessian with 1 negative eigenvalues
Warning messages:
1: In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
  unable to evaluate scaled gradient
2: In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
  Model failed to converge: degenerate Hessian with 1 negative eigenvalues
Warning messages:
1: In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
  unable to evaluate scaled gradient
2: In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
  Model failed to converge: degenerate Hessian with 1 negative eigenvalues
Warning messages:
1: In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
  unable to evaluate scaled gradient
2: In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
  Model failed to converge: degenerate Hessian with 1 negative eigenvalues
```

Therefore, I check the s_1_boot bootstrap object I created

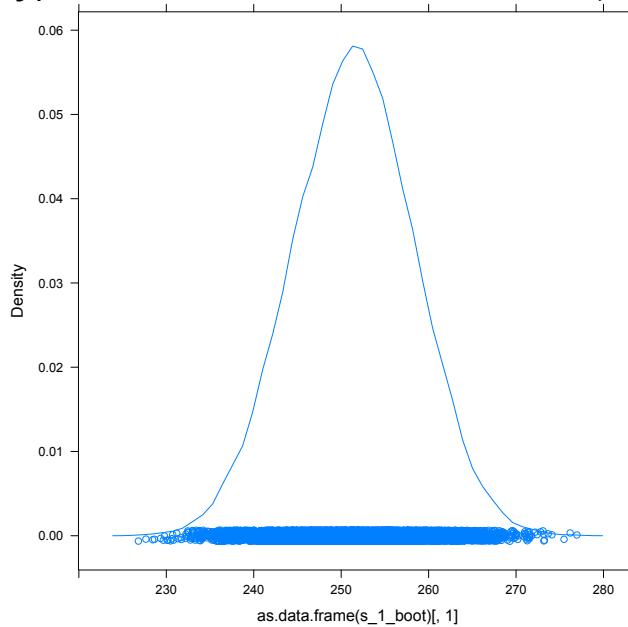
60

```
nrow(as.data.frame(s_1_boot)) # 10'000: good!
head(as.data.frame(s_1_boot))
tail(as.data.frame(s_1_boot))
as.data.frame(s_1_boot)

  (Intercept)      Days
1    260.9104  7.672451
2    248.5611 10.801059
3    243.8825 11.980033
4    249.8081 13.305948
5    255.1640  9.828370
6    254.2366 11.329536
7    258.3140  8.930460
8    252.5185 12.332080
9    263.3180  9.764524
10   253.4023  8.650229
11   248.5306 10.392461
12   257.6086 13.711539
13   242.5174 10.962705
etc...
```

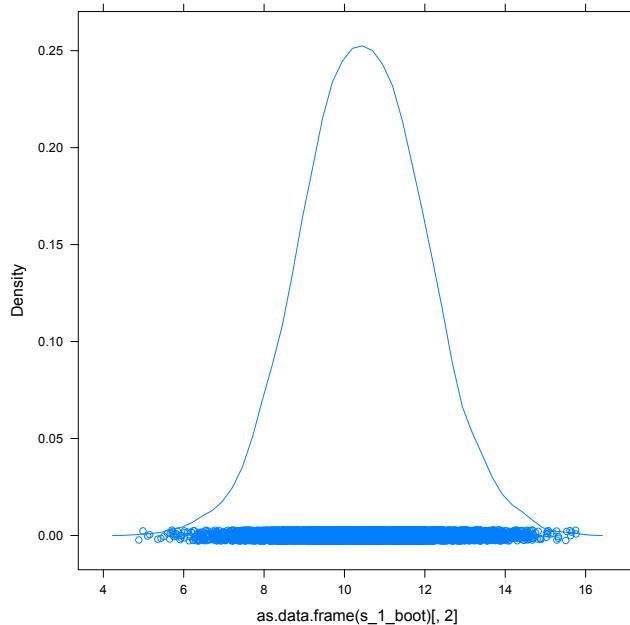
61

```
densityplot(as.data.frame(s_1_boot)[,1]
```



62

```
densityplot(as.data.frame(s_1_boot)[,2]
```



63

Level	Normal	Basic	Percentile
99.9%	(5.39, 15.55)	(5.55, 15.67)	(5.26, 15.39)

For comparison: Profile: (4.68, 16.26)

Bootstrapping: Days $p < .001$

We (again) conclude: There is a significant effect of days of sleep deprivation on reaction times.

64

How could you write this up?

Reaction times increased significantly as a function of the number of days with sleep deprivation
 (Estimate = 10.47(1.55), $p < .001$, 99.9% CI [5.26, 15.39].

Notes

- Estimate = regression coefficient (Std. Error) from summary output; you could also write coef = or B =
- Here, I report the percentile-based bootstrap CI (but profile or basic would also be fine)

65

Back to the formula

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

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

66

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

Extract fixed effects from model: \mathbf{b}_0 and \mathbf{b}_1

```
fixef(s_1)
(Intercept)      Days
 251.40510    10.46729
```

67

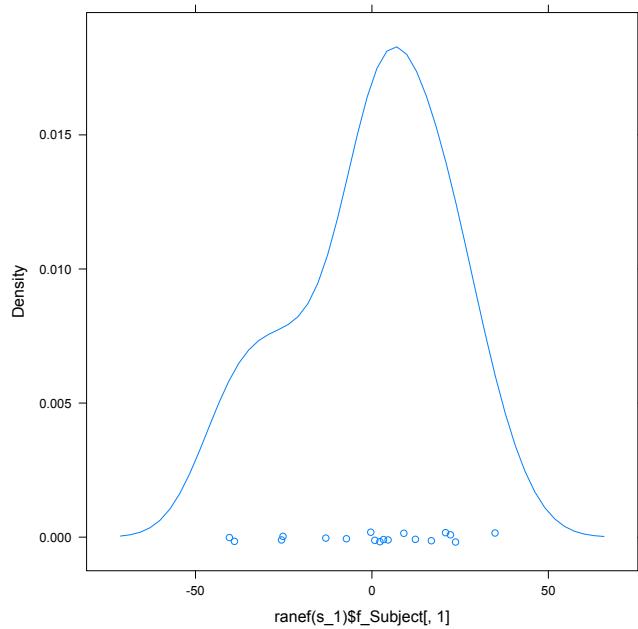
$$Y_{ij} = (b_0 + u_{0j}) + (b_1 + u_{1j})X_{ij} + \varepsilon_{ij}$$

Extract **random effects** from model: \mathbf{u}_{0j} and \mathbf{u}_{1j}

```
ranef(s_1)
$f_Subject
  (Intercept)      Days
pp_308   2.2585637  9.1989722
pp_309  -40.3985802 -8.6197026
pp_310  -38.9602496 -5.4488792
pp_330   23.6905025 -4.8143320
pp_331   22.2602062 -3.0698952
pp_332    9.0395271 -0.2721709
pp_333   16.8404333 -0.2236248
pp_334   -7.2325803  1.0745763
etc
```

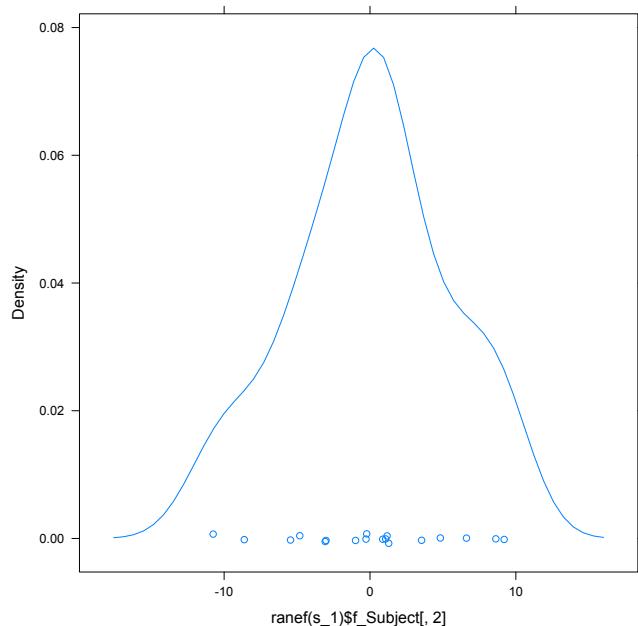
68

```
densityplot(ranef(s_1)$f_Subject[,1])
```



69

```
densityplot(ranef(s_1)$f_Subject[, 2])
```



70

Another nice way to plot the random effects

"Caterpillar plots"

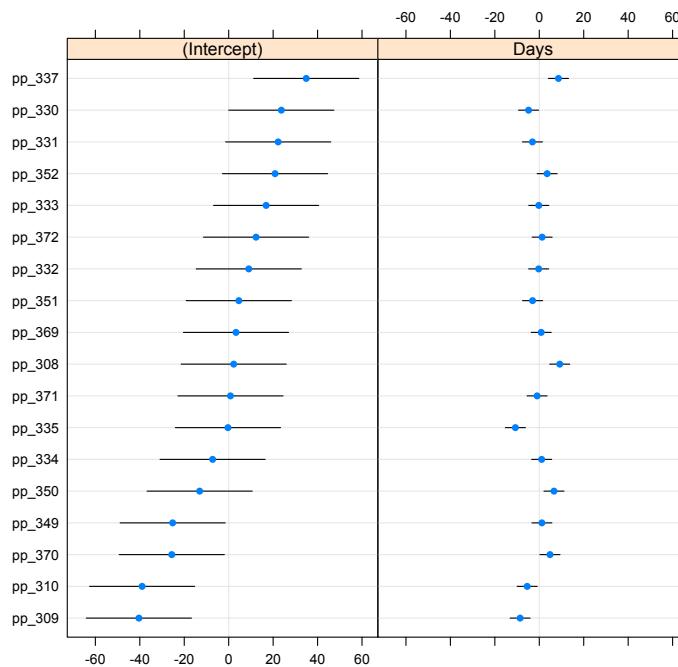
Show for each participant the estimated adjustment, plus an indicator for the uncertainty (95%)

Same x axis range for all random effects

```
dotplot(ranef(s_1, condVar = TRUE))
```

71

f_Subject

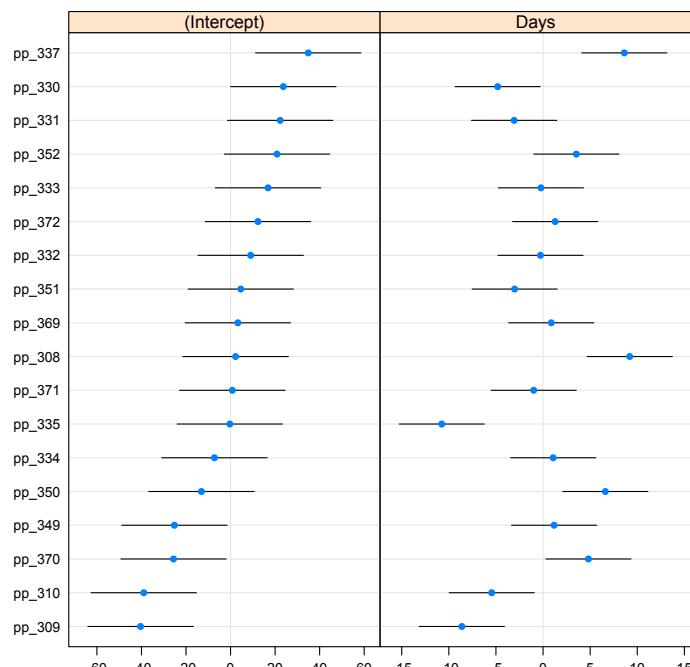


72

x axes adjusted

```
ranefs_s_1 <- ranef(s_1, condVar = TRUE)  
  
dotplot(ranefs_s_1, scales = list(x =  
  list(relation = 'free')))[["f_Subject"]]
```

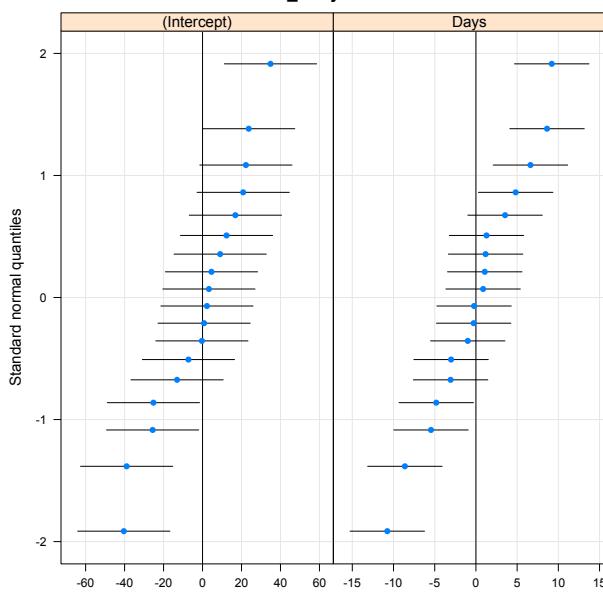
73

f_Subject

74

Variant: y axis = standard normal quantiles

```
qqmath(ranef(s_1, condVar = TRUE))
```



75

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

Extract BLUPs: (b_0+u_{0j}) and (b_1+u_{1j})

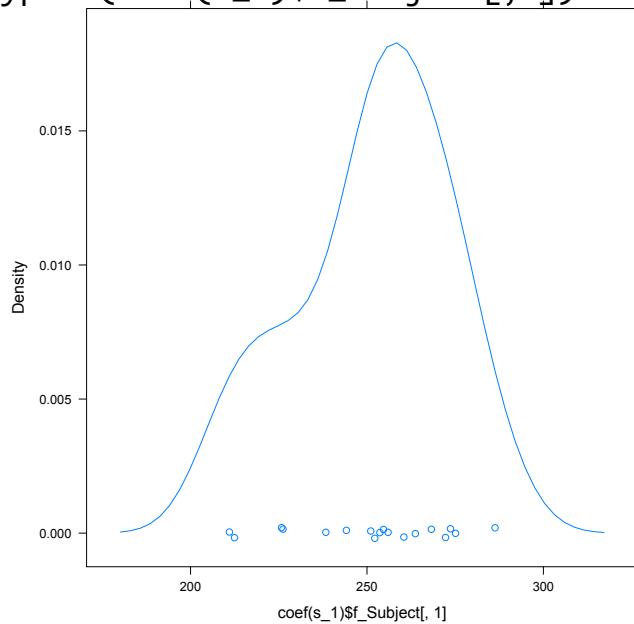
Best Linear Unbiased Prediction

http://en.wikipedia.org/wiki/Best_linear_unbiased_prediction

```
coef(s_1)
$f_Subject
  (Intercept)      Days
pp_308     253.6637 19.6662581
pp_309     211.0065  1.8475834
pp_310     212.4449  5.0184067
pp_330     275.0956  5.6529540
pp_331     273.6653  7.3973908
pp_332     260.4446 10.1951151
etc
```

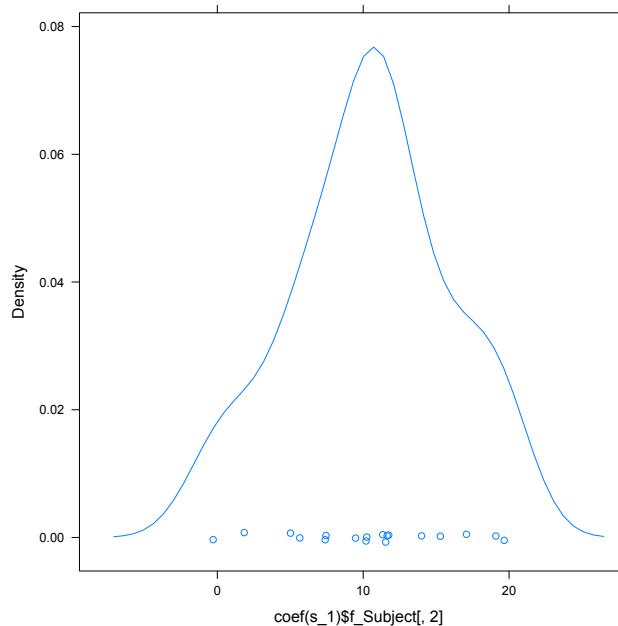
76

```
densityplot(coef(s_1)$f_Subject[,1])
```



77

```
densityplot(coef(s_1)$f_Subject[,2])
```



78

Today: More Theory and Models

- HW/recap/follow-up last week
- Random slopes
 - An example data set
 - Some theory
 - A model with random intercept and random slope
 - More plots
- Homework/lab session
 - My first interaction!

79

Today's Homework/Lab Session

Instructions: BlackBoard → Week 3

(1) Cherry Pit Data (same data as last week)

Answer the following research questions:

- Is there a significant trial effect?
- Is there a significant gender effect?
- Is there a significant gender x trial interaction?
- Are there significant individual differences in ...
 - ... how far participants can spit on average?
 - ... the trial effect?

→ Do all the relevant steps (including diagnostic plots, confidence intervals etc)

80

(2) The Sleep Study Data

- Not ALL the usual steps, but just some of them
- Try NOT to look at the code in my slides!

NOTE: Carefully read instructions, **you have to do some things differently than how I did them on the class slides!** (Which might lead to different results...)

81

What do you have to hand in?

- Answers to questions, output, figures, R code
- Hint: Figures can be decreased in size, so you end up with fewer pages to print out
- Paste everything into a word doc and
- Print it out, bring it to lab session on Feb 29

82

Questions or comments?

That's it for today's lecture!

See you in the basement at 13:45

SP A -1.55A+B