Linear Mixed-Effects Models

(aka Statistics III)

Bernd Figner
b.figner@psych.ru.nl

Week 7: March 24, 2014

1

Today: A mixed bag

- Some clarifications, rules of thumb, et al
- Finishing up multilevel perspective (FMF ch. 19)
  - growth-curve models
  - covariance structures
- Project presentations 1: Hannah's project
- Generalized mixed models: Overview
  - binary data
  - proportions ("summarized binary data")
  - counts
- glmer example: intertemporal choices

2

In-Class Exam

- Multiple-choice format
- Compared to take-home exam, stronger focus on more theoretical aspects
- But practical aspects also relevant (e.g., R syntax, reading/understanding outputs, etc)
- Relevant content: All materials
  - Classes: slides, example R scripts, ...
  - Mandatory literature/reading

3

Questions regarding home-work or other things?

4

Some Tips and Clarifications

Unload a specific library

```
detach("package:nlme", unload=TRUE)
```

Centering and scaling

(a) When I said **center** I meant subtracting the mean

```
scale(IV, center = TRUE, scale = FALSE)
```

(b) When I said **scale** I meant subtracting the mean AND dividing by the SD!

```
scale(IV, center = TRUE, scale = TRUE)
```

5

Rules of thumb

How complex can my model be?

- For each parameter in your model: at least 10-20 observations (rows of data)
- Also depends on how "informative" DV is (e.g., continuous vs. binary)

Random intercepts

- At least 5 different levels necessary, otherwise R cannot estimate variance properly
- What if < 5? (for example 4 different face stimuli)

→ **Add as fixed effect**

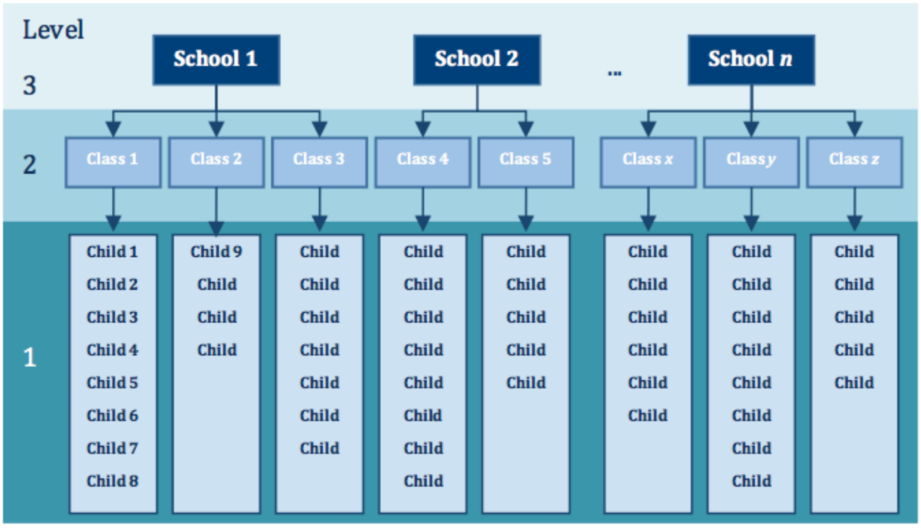
6

Multilevel Models

FMF book chapter 19

7

A Three-Level Hierarchy



Common Multilevel Procedure

First Step

- Determine dependency of observations in the data
  - "Baseline" model versus "null" model comparison
  - Intraclass correlation coefficient (ICC)
- Determine appropriate random effects structure
  - LRTs for random effects
  - Information criteria (e.g., AIC, BIC, DIC)

→ Data-driven model-selection stage

→ FMF: start with simple model, increase complexity

Second Step

- Use resulting model for inference about fixed effects

9

Two Views

Until last week

- Repeated measures and mixed-model perspective

→ Non-independence assumed based on theoretical reasons and/or study design

→ No reason to test it, we just model it

Multilevel perspective

- With nested/hierarchical data...
- ...dependence not always clear based on theoretical/ study-design reasons
- Use data to estimate (in)dependence

10

lme4 versus nlme

nlme is predecessor of lme4 (same developers)
- **lme4:** faster, more flexible, bootMer makes bootstrapping easy, ...
- **nlme:** `summary()` gives p values; different covariance structures to choose from

Recommendation: lme4, unless good reasons to use nlme
  - Specific covariance structure

11

Multilevel Models
FMF book chapter 19

Unfinished Business

→ Covariance Structures
→ Growth-curve models

12

Covariance Structures

nlme → possible to specify specific covariance structures

lme4 → not possible in the same way; often not necessary

Available covariance structures in nlme

- **Variance components (aka "independence model")**
  - Random effects are independent with similar variances
- **Diagonal**
  - Random effects are independent with different variances

  → same in lme4

  `(1|group) + (0+IV1|group) + (0+IV2|group)` etc

- **AR(1): First-Order Autoregressive**
  - Random effects are related with data points closer in time being more similar than those distant in time
  - Variances of random effects are similar

  → For example for longitudinal data

→ Not possible in lme4, but

- not necessarily a problem
- there might be other ways to achieve something similar (but not trivial in any case)

THUS: main reason to use nlme, if you need autoregressive covariance structures!

- **Unstructured**
  - Covariances and variances of random effects are "unpredictable"
  - I.e., they are estimated from the data

  → **That's what lme4 does by default and is also most common for models with random intercepts and slopes**

nlme has more (co)variance options (ARMA, ...)
→ **see `?corClasses` after loading library `nlme`**

To summarize

- nlme has more options to model specific (co)variance structures than lme4
- for most applications, however, lme4 does fine and has many advantages over nlme
- BUT: for data with autocorrelated residuals and no crossed random effects, it might be worth using nlme

Some related discussion about these issues:

http://stats.stackexchange.com/questions/5344/how-to-choose-nlme-or-lme4-r-library-for-mixed-effects-models

Questions so far?

Next: Growth-curve models

Growth-Curve Models

Nothing too special, really...

- Repeated-measures data
- The effect of time is of interest in itself
- **For example**
  - How does marriage satisfaction change over time? (FMF)
  - How does postural sway change over time? (Hannah)
  - ...
- Effects of time can be linear, quadratic, cubic (...)
- Same idea as our linear and quadratic predictors in the valuation ratings data! → use `poly()`

FMF Example: Honeymoon

Happiness as function of relationship (duration)

Example data in FMF book chapter: made up!

- Changes in life satisfaction as a function of duration of relationship → "honeymoon phase?"
- Life satisfaction at time 0, plus after 6, 12, and 18 months
- Some missing data (to demonstrate that mixed models can handle that)
- Data file: Honeymoon Period.dat → on BlackBoard
- Data from only 1 person per couple!

Data structure?

- Like repeated measures
- Each participant contributes 4 data points
- Life satisfaction (0 to 10): At 0, 6, 12, 18 months
- "Person" → pp code with numbers 1 to 123
- Gender: 0/1

Difference to repeated-measures data so far?

- Of main interest: time itself, i.e., changes over time

```
hm1 <- read.delim('Honeymoon Period.dat', header = TRUE)

> hm1[1:6, 1:3]
  Person Satisfaction_Base Satisfaction_6_Months
1      1                 6                     6
2      2                 7                     7
3      3                 4                     6
4      4                 6                     9
5      5                 6                     7
6      6                 5                    10
```

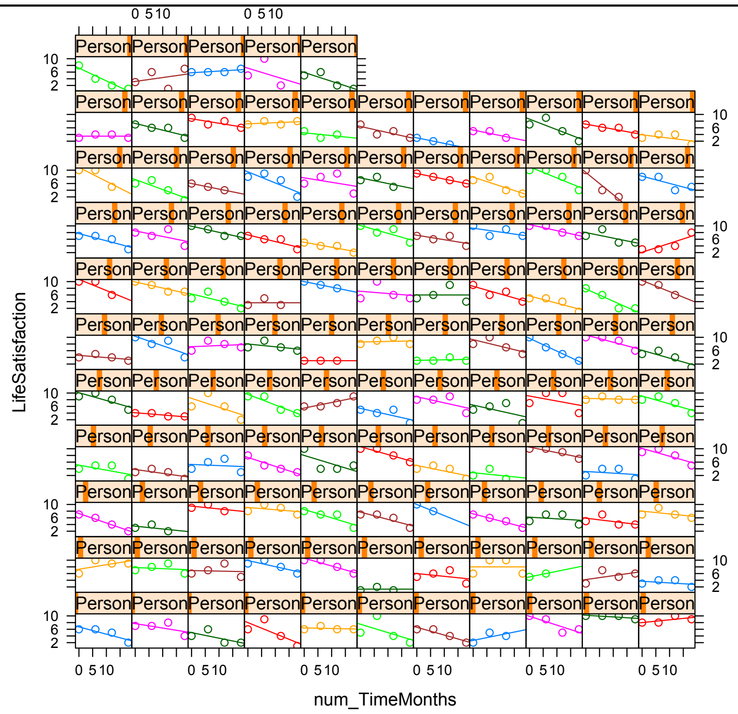
```
> hm1[1:6, 4:ncol(hm1)]

  Satisfaction_12_Months Satisfaction_18_Months Gender
1                      5                      2      0
2                      8                      4      1
3                      2                      2      1
4                      4                      1      0
5                      6                      6      0
6                      4                      2      1
```

Wide format!

- Thus, need to change to long format, e.g., using `melt()`
- Participant code → turn into explicit factor
- Turn Gender also into explicit factor

Then: check data, e.g., with `xyplot()`

- I first created an ordered factor with character entries for Time (whether it's baseline, 6, 12, or 18 months)
- Just for the plot, I also created a numeric variant (0, 6, 12, 18)



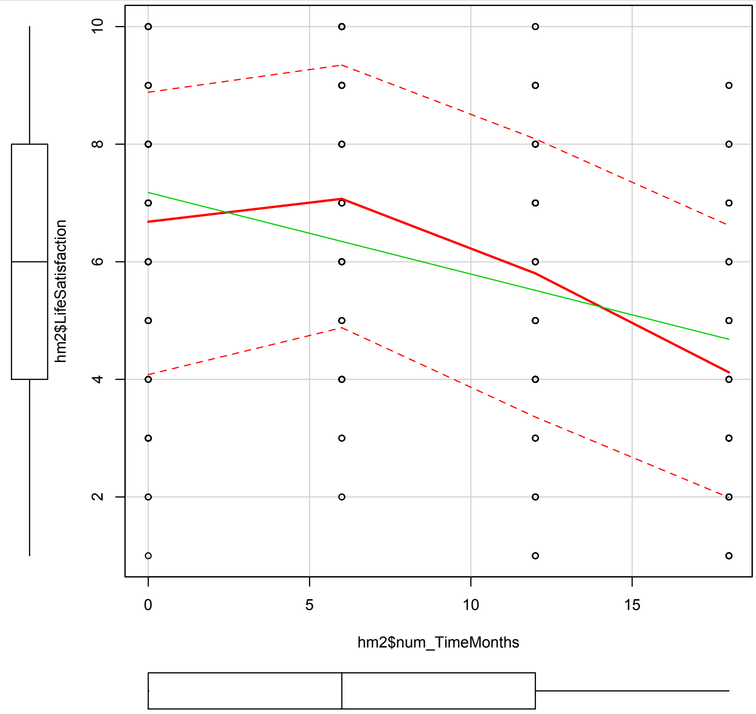
Evidence for non-linear trends?

`scatterplot()` from car

- Linear regression line + smoothed line (Loess)
- Add "jitter" to better see different data points

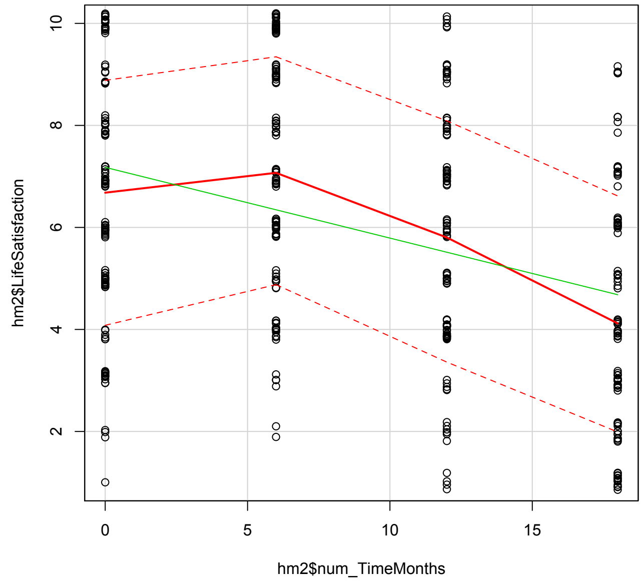
Without jitter

```
scatterplot(hm2$LifeSatisfaction ~
hm2$num_TimeMonths, smoother = loessLine)
```



With jitter, without boxplots

```
scatterplot(hm2$LifeSatisfaction ~ hm2$num_TimeMonths, smoother
= loessLine, jitter = list(x = 0, y = 1), boxplots = FALSE)
```



FMF Models: nlme

- Using first-order autoregressive covariance structure
- Assumption: The closer 2 data points are in time, the more highly correlated
- They go again through all the steps starting with `gls()`
- Non-standard optimizer, as otherwise some models don't converge

NOTE: I got an error when I try to run their final model

```
Error in structure(res, levels = lv, names = nm, class = "factor") :
  'names' attribute [323] must be the same length as the vector [0]
```

→ After turning Time variable into explicitly numeric variable (0, 6, 12, 18), it worked!

Most complex FMF model

Time: linear, quadratic, cubic

```
polyModel<-lme(Life_Satisfaction ~ poly(Time, 3), data =
restructuredData, random = ~Time|Person, correlation =
corAR1(0, form = ~Time|Person), method = "ML", na.action
= na.exclude, control = list(opt="optim"))
```

NOTES

- Time linear and quadratic significant, cubic not significant
- Random slope only for linear, but not quadratic and cubic Time effects!

In lme4

- Cannot use autoregressive covariance structure
- Otherwise equivalent model to FMF
  - create first 3 separate predictors for Time linear, quadratic, and cubic using `poly()` → to be able to use `bootMer()` or `Anova()` for p values

```
fmf_poly_lme4 <- lmer(LifeSatisfaction ~
Time_lin + Time_quad + Time_cub + (1 + Time_lin
| Person), data = hm2,  REML = FALSE)
```

Equivalent results

- Time linear and quadratic significant, cubic not

Fully Maximal Model?

- Random slopes for linear, quadratic, and cubic Time?
- **Not possible...**

```
number of observations (=438) <= number of random effects
(=460) for term (1 + Time_lin + Time_quad + Time_cub |
Person); the random-effects parameters and the residual
variance (or scale parameter) are probably unidentifiable
```

Too complex for data

→ Same case as with 1 observation per factor combination (4 data points per participant!)

→ Need to simplify, e.g., remove cubic effect; ...

→ see Barr et al. advice and my BB example R script

End of multi-level part

FMF book chapter: 2 more data sets

- "Labcoat Leni's Real Research:" Lap dancer data
  → Miller et al. (2007).dat
- "Smart Alex's tasks:" Exercise in children
  → Hill et al. (2007).dat → Nice example!

Data are on BlackBoard; as are example scripts how I would analyze those data

- Honeymoon data and my respective example R script: also on BlackBoard

Questions? Comments?

Hannah's Presentation

therafter:
Generalized linear mixed models (GLMMs)
BOLKER/Jaeger

34

Generalized Linear Mixed Models `glmer()`

35

- **So far: lmer → DV**
  - Continuous (more or less)
  - Normally distributed
- **Not all our DVs are like that:**
  - **Binary** data (e.g., SS vs. LL in intertemp. choice) → logistic models
  - **Proportions** (e.g., proportion correct responses); summarized binary responses → binomial models
  - **Counts** (how often does something occur?) → often Poisson models

36

Generalized Linear Mixed Models

Non-normal, i.e., categorical data

- Binary (0/1) → logistic → family = binomial
- Proportions (= summed 0/1) → family = binomial
- Counts → family = Poisson
- [More than 2 categories → multinomial mixed models]

37

ANOVA with such data?

- Jaeger (2008): VERY BAD IDEA!!
- yes, even after transformation

Sometimes a bit of a grey zone

- DV: very normal and continuous; no "stacking up" of values at either end of scale

→ Treating it as continuous often ok

→ Try gaussian and generalized model → see which model better predicts data (e.g., plot DV versus fitted)

38

Bolker et al. (2009)

→ **Bolker page: http://glmm.wikidot.com/ and http://glmm.wikidot.com/faq**

→ Not the easiest paper, but touches upon several important aspects

- Ecology/evolution → some translations
  - "Block" → grouping variable (for us: participants)
  - "Covariate" → continuous predictor
  - "Sample size" → number of observations (number of rows in data frame)
- Highlights how challenging GLMMs can be...

39

Challenging GLMM World...

- **Different estimation procedures**
  - Pseudo-/Penalized Quasilikelihood
  - Laplace approximation
  - Gauss-Hermite quadrature
- **Different ways to get p values**
  - Wald Z, Wald Chisquare, Wald t, LRT
  - plus Bootstrapping: `bootMer()` and `PBmodcomp()`
- **GLMMs = active area of research**
  - Few established guidelines (e.g., model diagnostics)
  - Less ready-made commands than for other models

40

Jaeger (2008)

- Using ANOVA for proportions (or even binary responses): still widespread (still today!)
- Can lead to misleading results even after transformations
- Independent observations → generalized linear models: `glm()`
- "Grouped errors" → generalized linear mixed-effects models: `glmer()` in lme4

41

glmer()

- **How does lme4 know that it should do a *generalized* model?**

```
glmer(DV ~ IV1 + IV2+ (1 + IV1|group),
family=binomial, data = …, control =
glmerControl(...))
```

Different families, most typical ones:

- logistic and proportions → binomial
- Counts → poisson

42

glmer()

- **You can also specify different link functions → same as for `glm()`**

```
glmer(DV ~ IV1 + IV2+ (1 + IV1|group),
family=binomial(link = "logit"), data =
…, control = glmerControl(...))
```

43

Non-convergence quite frequent...

- **Increasing number of iterations**

```
glmer(..., control = glmerControl(optCtrl =
list(maxfun = 1000000))
```

- **Try different optimizers**
  - bobyqa
  - Nelder_Mead
  - optimx from package optimx

44

- **bobyqa**

```
glmer(..., control = glmerControl(optCtrl
= list(maxfun = 1000000), optimizer =
"bobyqa"))
```

- **Nelder_Mead**

```
glmer(..., control = glmerControl(optCtrl
= list(maxfun = 1000000), optimizer =
"Nelder_Mead"))
```

- **Nelder_Mead AND bobyqa**

```
glmer(..., control = glmerControl(optCtrl
= list(maxfun = 1000000), optimizer =
c("Nelder_Mead", "bobyqa"))
```

45

optimx: a bit more complicated

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

- **Without increasing number of iterations**

```
glmer(..., control =
glmerControl(optimizer = 'optimx',
optCtrl = list(method = 'nlminb')))
```

- **WITH increasing number of iterations**

```
glmer(..., control =
glmerControl(optimizer = 'optimx',
optCtrl = list(method = 'nlminb', maxit =
10000)))
```

46

Logistic GLMMs

Binary DV (0/1) → „logit transformed"

- http://en.wikipedia.org/wiki/Logit
- Values between 0 and 1 → values from large negative, over zero (logit of 0.5 is 0), to large positive
- Model computations on „logit scale"

Remember (from `glm()`)

- Model output: on „logit scale" (e.g., the intercept)
- Back-transform to the „probability scale" (0 to 1):
  → `plogis()`

NOTE: `fitted(model)` → on „probability scale"

47

Logistic models: „Less information"

- Binary → less info; compared to continuous DV
- Given same number of observations:
- Less parameters can be fit (compared to lmer)

→ More likely to get non-convergence

→ More often necessary to simplify models

- Barr et al. (2013) steps
- More substantial simplification: less fixed-effects (main effects, interactions), ...

48

How to get p values?

- `summary(model)` gives p values → hooray!
- Do NOT use the p values from the summary statement (= „Wald Z tests“) → not precise
- Alternatives
  - `bootMer()` or `PBmodcomp()`
  - LRTs using `anova()` or `drop1()`
  - Conditional F tests? nope, only for gaussian models → ~~`Anova(..., test = 'F')`~~; ~~`KRmodcomp()`~~
- Recommendation from worst to best: http://glmm.wikidot.com/faq

49

Example Data

The intertemporal choice task

NOW Trial

SS = Sooner Smaller LL = Later Larger

SS	LL
€ 40.60	**€ 45.70**
Today	**4 Weeks**

51

NOT-NOW Trial

SS = Sooner Smaller LL = Later Larger

SS	LL
€ 40.60	**€ 45.70**
2 Weeks	**6 Weeks**

52

Task Design

1. Now versus not-now trials

- **Now trials:** Immediate SS vs. future LL
  - today vs. 3 days
  - today vs. 14 days
  - today vs. 28 days
- **Notnow trials:** Future SS vs. future LL
  - 14 vs. 17 days
  - 14 vs. 28 days
  - 14 vs. 42 days

More impatient if immediate rewards available?

present-bias aka immediacy effect aka now effect

53

2. Time Difference (LL – SS)

- 3 days
- 14 days
- 28 days

More impatient if one has to wait longer?

54

3. Relative differences in amounts

5 different levels of differences

- ~ -20% (catch trials: SS > LL!)
- ~ 5%
- ~ 10%
- ~ 20%
- ~ 30%
- ~ 50%

→**Exact relative differences vary around these categories** („jitter“ to reduce memory effects; rounding errors)

→Categories for graphs; exact differences for models

More patient if one gets more for waiting?

55

4. SS Amount

- Range: € 16 to € 70
- (pseudo)randomly drawn from normal distribution
- not factorially varied
- not of interest here; varied to avoid memory effects and discourage simple rule-based choice strategy
- should be included in model (nuisance variable)
- "magnitude effect:" more money → more patience

56

Files on BB

Data

- Raw data: ugly → needs preprocessing
- Preprocessed data → SS choice = 0; LL = 1
  - catch trials removed

R scripts

- pre-processing script
- script for figures and glmer models

57

Some Figures

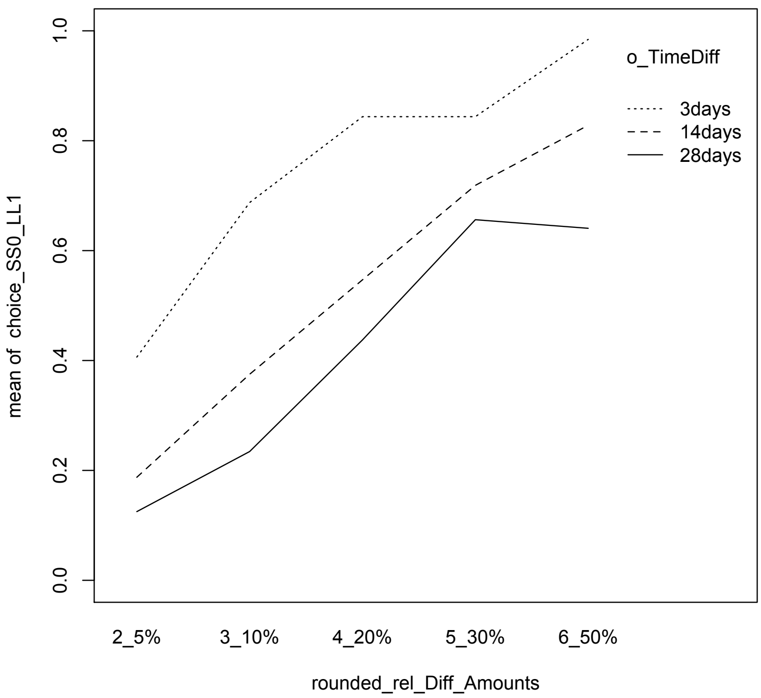
DV: 0 = SS choice; 1 = LL choice

→Mean = 'proportion of LL choice"

(1) LL choice as function of relative amount difference and time difference

```
with(itc4, interaction.plot(rounded_rel_Diff_Amounts,
o_TimeDiff, choice_SS0_LL1))
```

58



59

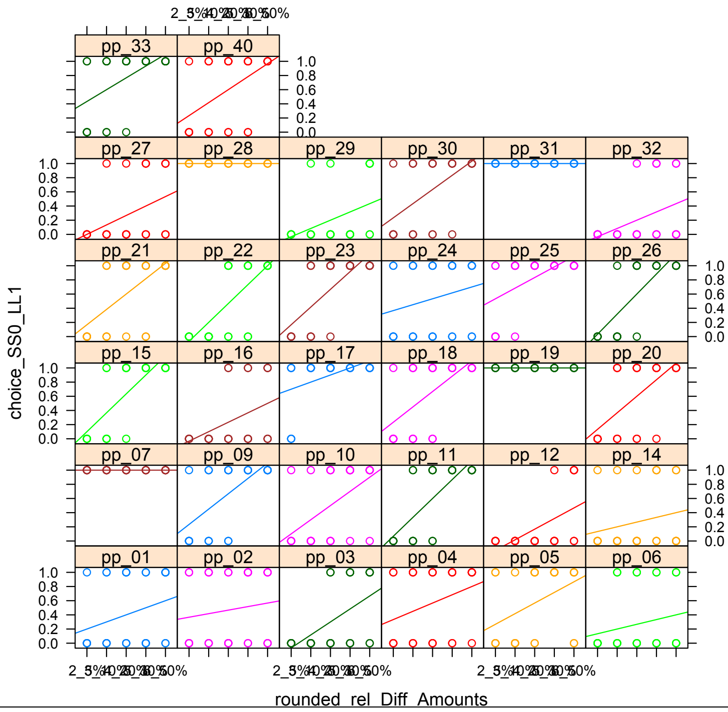
Show pp variability?

(1) LL choice as function of relative amount difference and participant

xyplot()

```
xyplot(choice_SS0_LL1 ~ rounded_rel_Diff_Amounts |
pp_code_ITC, groups = pp_code_ITC, data = itc4, type =
c('p', 'r'), auto.key = FALSE)
```

60

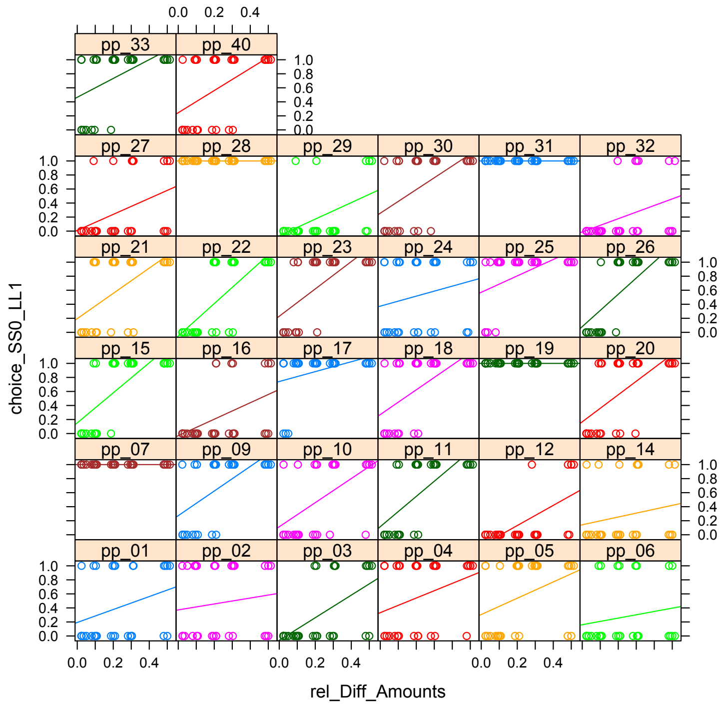


61

Use exact relative differences

```
xyplot(choice_SS0_LL1 ~ rel_Diff_Amounts | pp_code_ITC,
groups = pp_code_ITC, data = itc4, type = c('p', 'r'),
auto.key = FALSE)
```

62



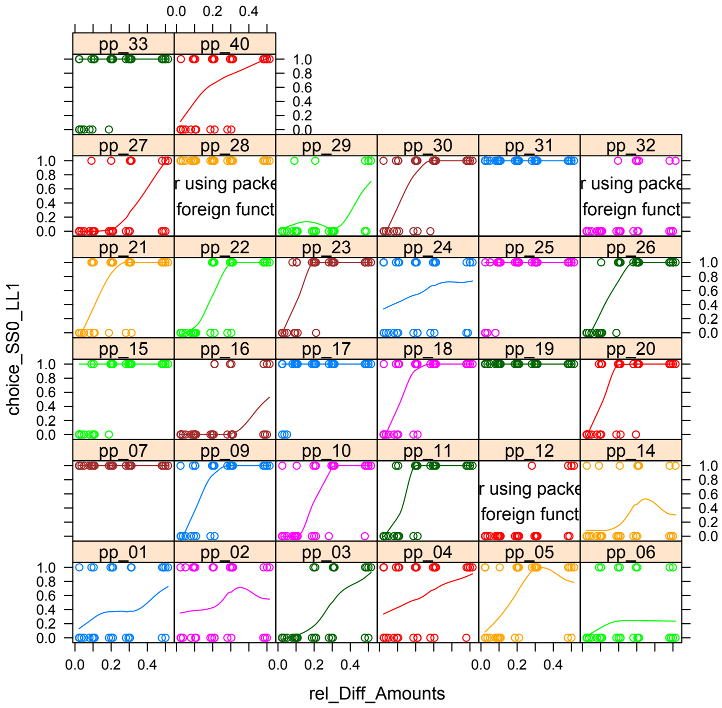
63

Shows linear relationship in "probability space" (not logit space)

```
xyplot(choice_SS0_LL1 ~ rel_Diff_Amounts | pp_code_ITC,
groups = pp_code_ITC, data = itc4, type = c('p',
'smooth'), auto.key = FALSE)
```

Not ideal, but a quick'n'dirty solution...

64



65

Some Models

Choice as a function of:

- Now/notnow
- Time difference
- Relative Amount difference
- SS Amount (magnitude effect)

→ Now/notnow ordered factor; all other continuous (center/scale!)

→All within-subject → random slopes

66

```
m2_1 <- glmer(choice_SS0_LL1 ~ o_Now_Notnow +
s_TimeDiff + s_SS_Amount + s_rel_Diff_Amounts +
(1 + o_Now_Notnow + s_TimeDiff + s_SS_Amount +
s_rel_Diff_Amounts | f_pp_code_ITC),
family=binomial(link = "logit"), data = itc4)
```

→ **NOT converged!**

→ more iterations (10,000,000): still not converged...

→ let's have a look at summary()

67

```
Random effects:
 Groups        Name               Variance Std.Dev. Corr
 f_pp_code_ITC (Intercept)        9.81445  3.1328
               o_Now_Notnow.L     0.04851  0.2203    1.00
               s_TimeDiff         0.58522  0.7650    0.43  0.46
               s_SS_Amount        0.07150  0.2674    0.20  0.20 -0.38
               s_rel_Diff_Amounts 1.58327  1.2583    0.67  0.68  0.26  0.77
Number of obs: 960, groups: f_pp_code_ITC, 32

Fixed effects:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)          1.4077     0.6284   2.240   0.0251 *
o_Now_Notnow.L       0.1519     0.1759   0.864   0.3879
s_TimeDiff          -1.3542     0.2088  -6.487 8.78e-11 ***
s_SS_Amount          0.8760     0.1554   5.636 1.74e-08 ***
s_rel_Diff_Amounts   2.6125     0.3772   6.926 4.34e-12 ***
```

68

Next Steps

- **Removed now/notnow random slope**
  - still not converged
- **Removed all random covariance terms**
  - still not converged
- **Different optimizer (bobyqa)**
  - hooray!!
- **Go back to first model, using bobyqa**
  - no convergence
- **Remove now/notnow random slope**
  - hooray!! → my final model

69

Diagnostic Plots?

- **More complicated than for lmer models**
- **For today, some links for the interested**
- http://stats.stackexchange.com/questions/89991/use-predicted-values-with-or-without-random-part-to-plot-residuals-with-binnedpl
- http://stats.stackexchange.com/questions/63566/unexpected-residuals-plot-of-mixed-linear-model-using-lmer-lme4-package-in-r
- http://stats.stackexchange.com/questions/70783/how-to-assess-the-fit-of-a-binomial-glmm-fitted-with-lme4-1-0
- http://www.r-bloggers.com/model-validation-interpreting-residual-plots/
- http://www.sagepub.com/upm-data/38503_Chapter6.pdf

70

Getting p values

- `Anova(..., test = 'F')`
  - nope, only for gaussian models
- `KRmodcomp()`
  - nope, only for gaussian models
- **Possible options**
  - `bootMer()`
  - `PBmodcomp()`
  - `drop1()` or `anova()`

Results

All approaches lead to same conclusions
(although p values differ somewhat)

- Now/notnow: not significant
- Time difference: significant
- Relative Amount difference: significant
- SS Amount: significant

On BlackBoard

Course Documents→Data from lab session→Intertemp Choice: Choice Data

Raw Data

- **Data:** `ICT_MixMod_Choices_numlet.csv`
- **Pre-processing script:** `ITC_Preprocessing_GIF_SS_LL_Dutch_22March2014.R`

Preprocessed Data and script for models

- **Data:** `ICT_stacked_reduced_forBB_22March2014a.csv`
- **Figures and glmer models:** `MixMod_ICT_GLMER_ExampleScript_23March2014a.R`

73

Not Homework, but...

BlackBoard: Data sets and example scripts

FMF book

- Honeymoon data
- Lap dancer data: Miller et al. (2007).dat
- Exercise in children: Hill et al. (2007).dat → Nice!

Hannah's data: Body sway (reduced data set!)

Our own lab data

- Intertemporal choices: raw and preprocessed, with respective R scripts

- Use them to practice and get more experience!
- If possible, use example scripts only as backup if you get stuck or to compare when you are finished

74

Basement???

Good luck and have fun with the Take-Home Exam!!