Linear Mixed-effects Regression Model (LMER)

Yun (Renee) Zhang, PhD

Email: zhangy@jcvi.org

Monthly Biostatistics Discussion

10-17-2018

Outline

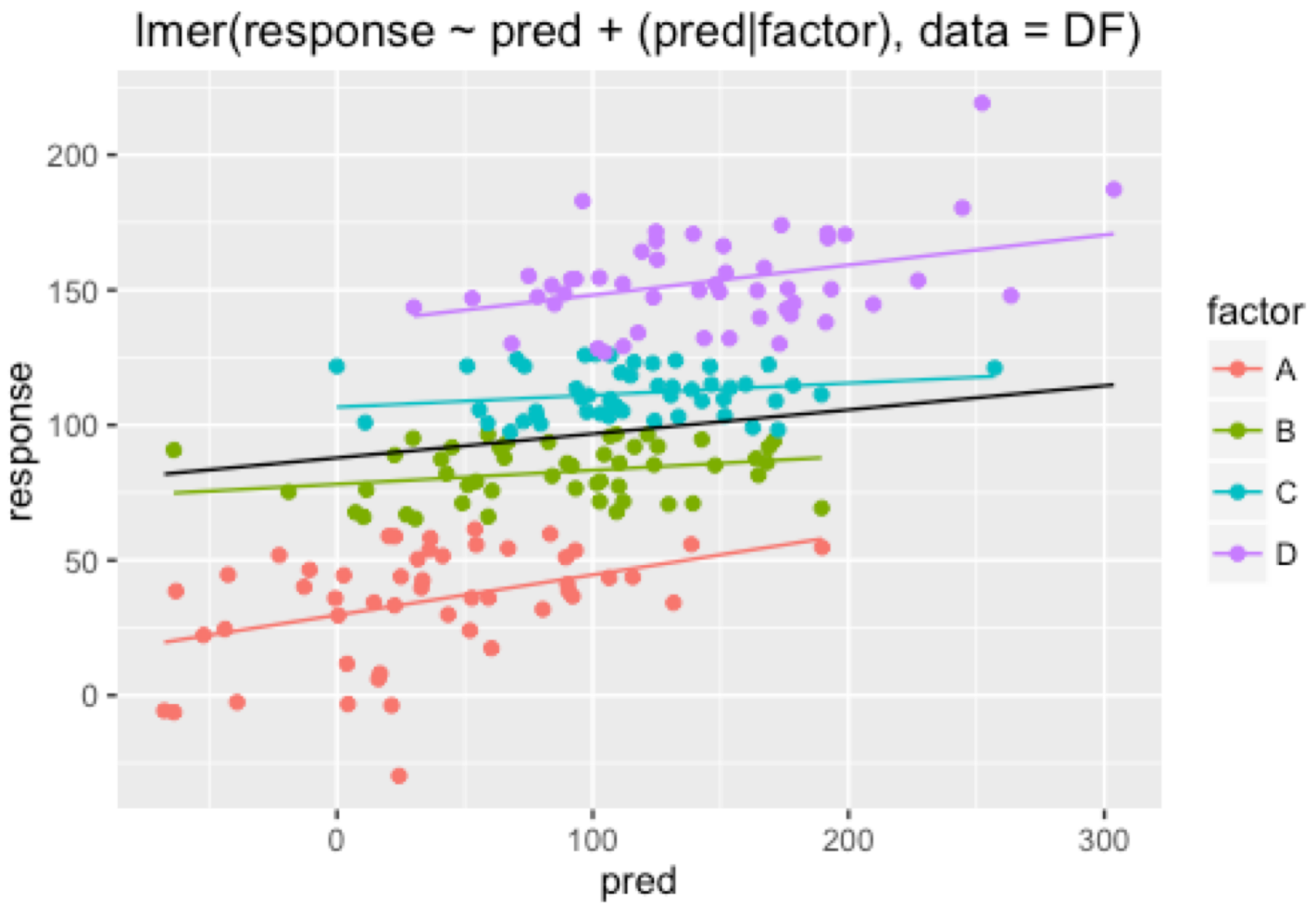
1. Linear Mixed-effects Model
2. Example 1: Repeated Measures ANOVA
3. Example 2: Longitudinal Study

Linear Mixed-effects Model

What Are Mixed-effects Models?

- Linear models
  - Observations assumed independent
  - Effects described as fixed effects
  - Errors are uncorrelated
  - Outcome = Intercept + (Fixed Effects) + Error
- Mixed-effects models
  - Observations not assumed independent
  - Random effects fitted to model additional source of variation
  - Errors may be correlated
  - Outcome = Intercept + (Fixed Effects) + (Random Effects) + Error
- Repeated measures nested within individuals (e.g., longitudinal studies)
- Groups nested within super-groups (e.g., cities nested within countries, clustered data)

Graphical Intuition



Model Specification

$$\boldsymbol{y} = X\boldsymbol{\beta} + Z\boldsymbol{\gamma} + \boldsymbol{\epsilon}$$

Fixed effects Random effects

- $\boldsymbol{\gamma} \sim N(\mathbf{0}, \Sigma_{\gamma}), \boldsymbol{\epsilon} \sim N(\mathbf{0}, \Sigma_{\epsilon})$
- Fixed effects: one or more predictors, e.g.

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2$$

```
response ~ predictor1 + predictor2
```

- Random effects: factor variables
  - Random intercept

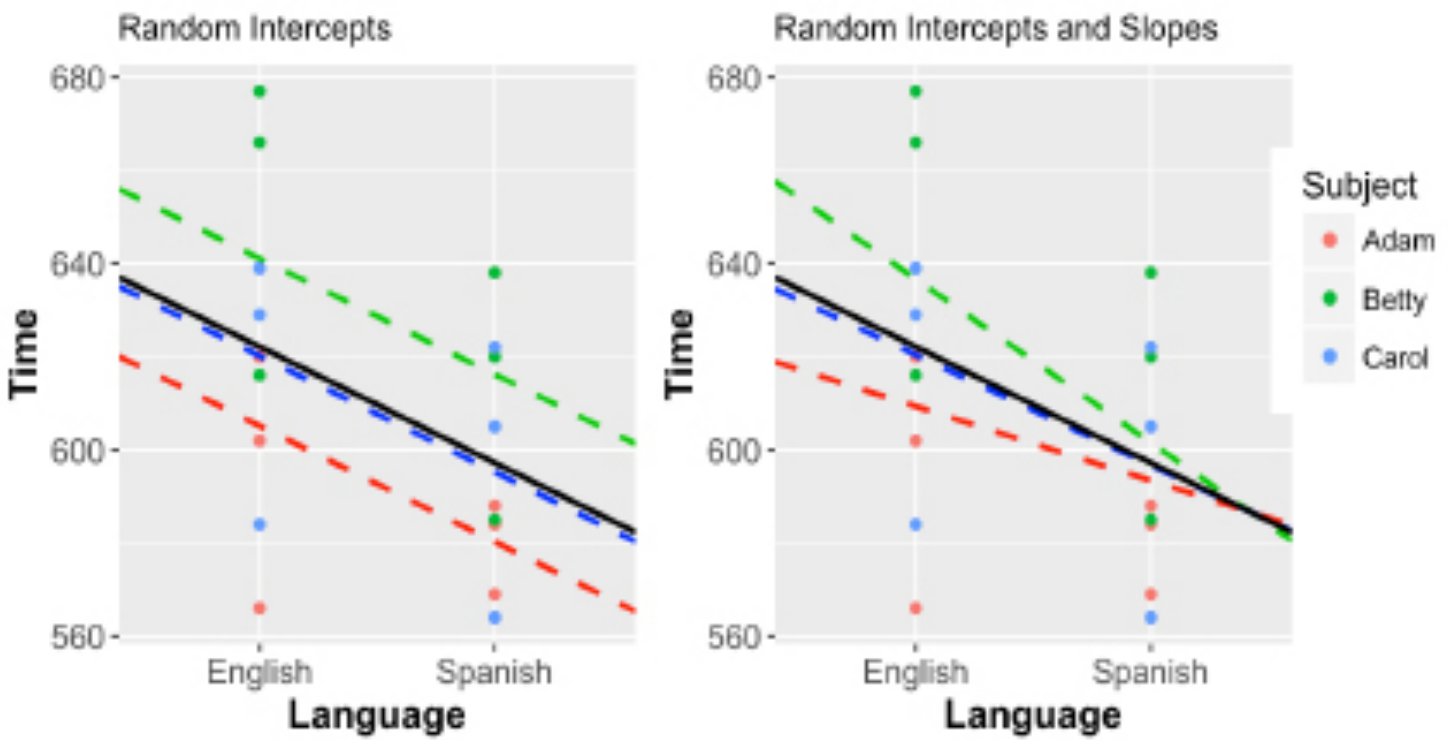
```
(1|factor)
```

  - Random intercept and random slope

```
(predictor|factor)
```

Random Effects

- Random intercept and random slope



Multilevel Model

- Level-1:

$$y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij}$$

- Level-2.1:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j}$$

- Level-2.2:

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + u_{1j}$$

- Overall:

$$y_{ij} = \underbrace{[\gamma_{00} + \gamma_{01}Z_j}_{\text{Intercept}} + \underbrace{(\gamma_{10}+\gamma_{11}Z_j)X_{ij}]}_{\text{Slope}} + \underbrace{[u_{0j} + u_{1j}X_{ij} + r_{ij}]}_{\text{Error}}$$

$$y_{ij} = \underbrace{[\gamma_{00} + \gamma_{10}X_{ij}}_{\text{Fixed effects}} + \underbrace{(\gamma_{01} + \gamma_{11}X_{ij})Z_j]}_{\text{Random effects}} + \underbrace{[u_{0j} + u_{1j}X_{ij} + r_{ij}]}_{\text{Error}}$$

R Packages

- `lme4`: Bates, D., Mächler, M., Bolker, B. and Walker, S., 2014. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.
  - Cited by 11127 according to Google Scholar
  - Philosophically, does not provide p-values
- `lmerTest`: Kuznetsova, A., Brockhoff, P.B. and Christensen, R.H.B., 2017. lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13).
  - Cited by 2486 according to Google Scholar
  - Practically, provides p-values with adjusted degrees of freedom

Example 1: Repeated Measures ANOVA

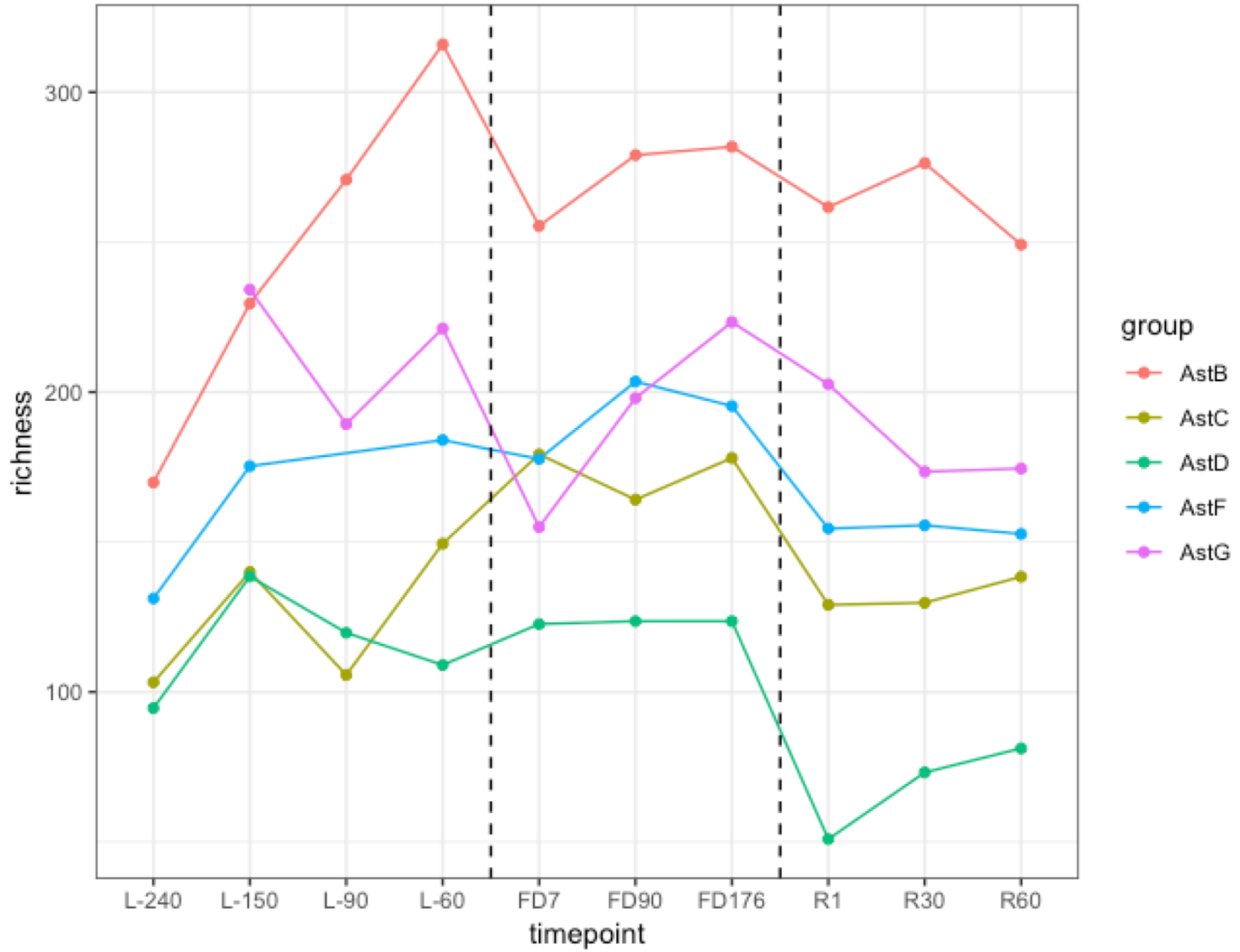
Thanks to Dr. Hernan Lorenzi

Flight Data

	group	flight	timepoint	richness	shannon
AstB_st_01	AstB	pre	L-240	169.85523	2.917034
AstB_st_02	AstB	pre	L-150	229.49118	3.861971
AstB_st_03	AstB	pre	L-90	270.80329	4.151797
AstB_st_04	AstB	pre	L-60	315.91470	4.238076
AstB_st_05	AstB	in	FD7	255.37437	3.246808
AstB_st_06	AstB	in	FD90	279.00000	4.187039
AstB_st_07	AstB	in	FD176	281.78161	3.483803
AstB_st_08	AstB	post	R1	261.64464	4.062190
AstB_st_09	AstB	post	R30	276.34073	3.802455
AstB_st_10	AstB	post	R60	249.16534	4.116311
AstC_st_01	AstC	pre	L-240	103.23431	2.585520
AstC_st_02	AstC	pre	L-150	140.00568	2.267304
AstC_st_03	AstC	pre	L-90	105.66620	1.811256
AstC_st_04	AstC	pre	L-60	149.44578	2.854151
AstC_st_05	AstC	in	FD7	179.25181	3.702806
AstC_st_06	AstC	in	FD90	164.12839	3.370611
AstC_st_07	AstC	in	FD176	178.00000	3.491467
AstC_st_08	AstC	post	R1	129.04502	2.590821
AstC_st_09	AstC	post	R30	129.74536	2.239643
AstC_st_10	AstC	post	R60	138.43668	2.686230
AstD_st_01	AstD	pre	L-240	94.63196	3.174327
AstD_st_02	AstD	pre	L-150	138.52526	2.940120

Species Richness

- Hypothesis: if species richness changes inflight or postflight with respect to the preflight baseline values



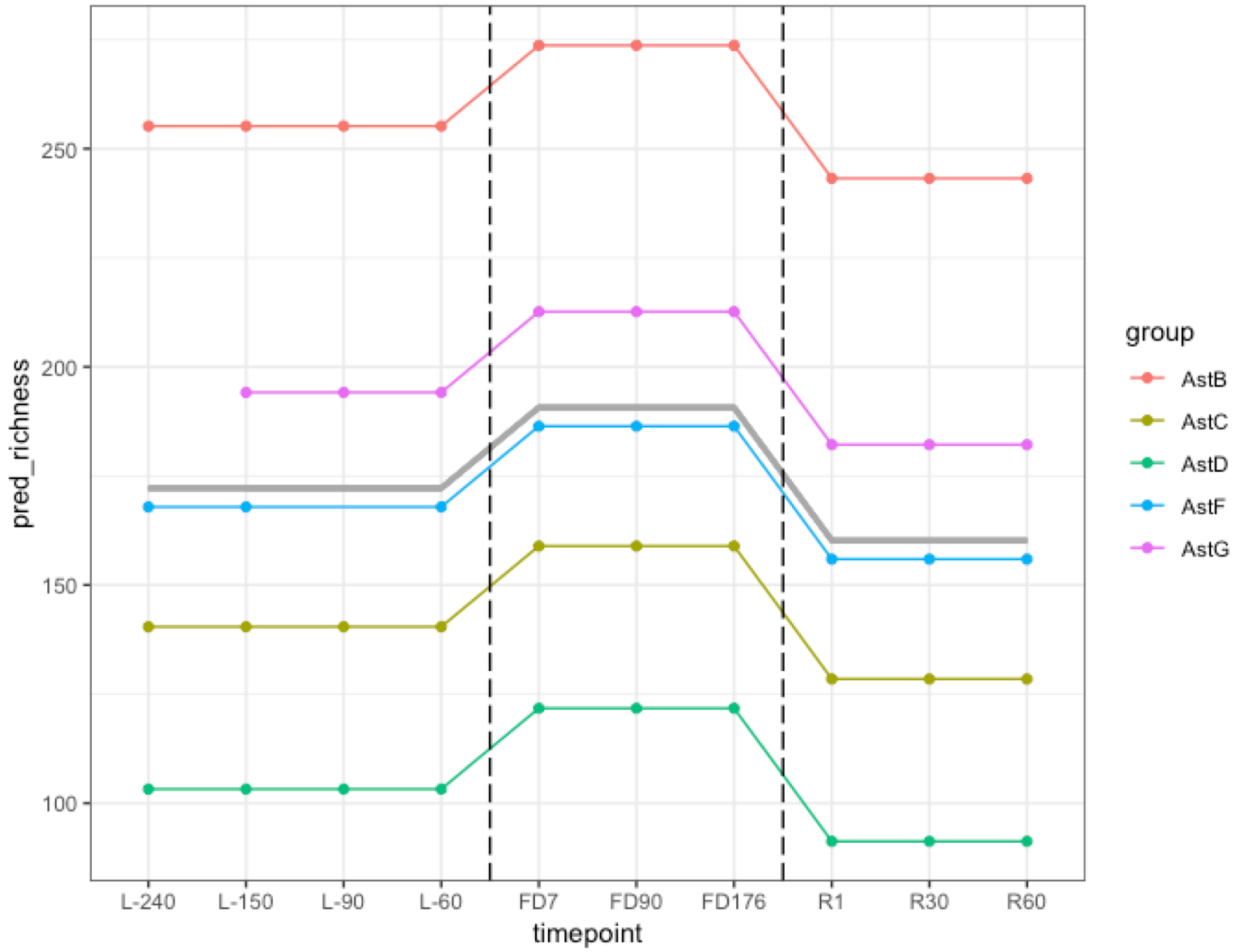
Mixed-effects Model 1: Random Intercept

Model specification

```
> mod.richness <- lmer(richness ~ flight + (1|group), data = dat)
> summary(mod.richness)
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: richness ~ flight + (1 | group)
   Data: dat

REML criterion at convergence: 447.6

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.1727 -0.3242  0.0693  0.5850  2.2576

Random effects:
 Groups   Name        Variance Std.Dev.
 group    (Intercept) 3362.6   57.99
 Residual              723.5   26.90
Number of obs: 48, groups:  group, 5

Fixed effects:
            Estimate Std. Error      df t value Pr(>|t|)
(Intercept)  172.176     26.702   4.307   6.448   0.0023 **
flightin      18.507      9.418  41.013   1.965   0.0562 .
flightpost   -11.974      9.418  41.013  -1.271   0.2108
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

p-value

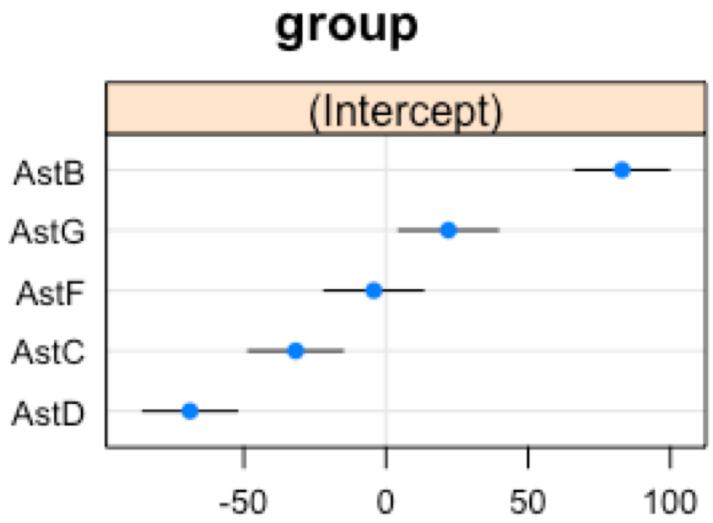
Fitted Values



Prediction of Random Effects

- Fixed effects *estimators*
- Random effects *predictors*

```
> ranef(mod.richness)
$group
     (Intercept)
AstB   83.015398
AstC  -31.756434
AstD  -68.946159
AstF   -4.277298
AstG   21.964493
```



- Hard to make inference on random effects (i.e. no p-values)

Compare with Linear Model

```
> lm.richness <- lm(richness ~ flight, data = dat)
> summary(lm.richness)

Call:
lm(formula = richness ~ flight, data = dat)

Residuals:
     Min       1Q   Median       3Q      Max
-109.247  -36.771   -6.609   35.089  144.722

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   171.19      14.33  11.944  1.5e-15 ***
flightin       19.49      21.26   0.917    0.364
flightpost    -10.99      21.26  -0.517    0.608
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- Similar estimates for fixed effects, but often larger standard errors
- Significance might be washed away

Repeated Measures ANOVA

- Use random effects to capture subject-specific correlation
- Significant using mixed-effects model

```
> anova(mod.richness)
Type III Analysis of Variance Table with Satterthwaite's method
       Sum Sq Mean Sq NumDF  DenDF F value  Pr(>F)
flight 7087.6  3543.8     2 41.011  4.8983 0.01237 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- Not significant using linear model

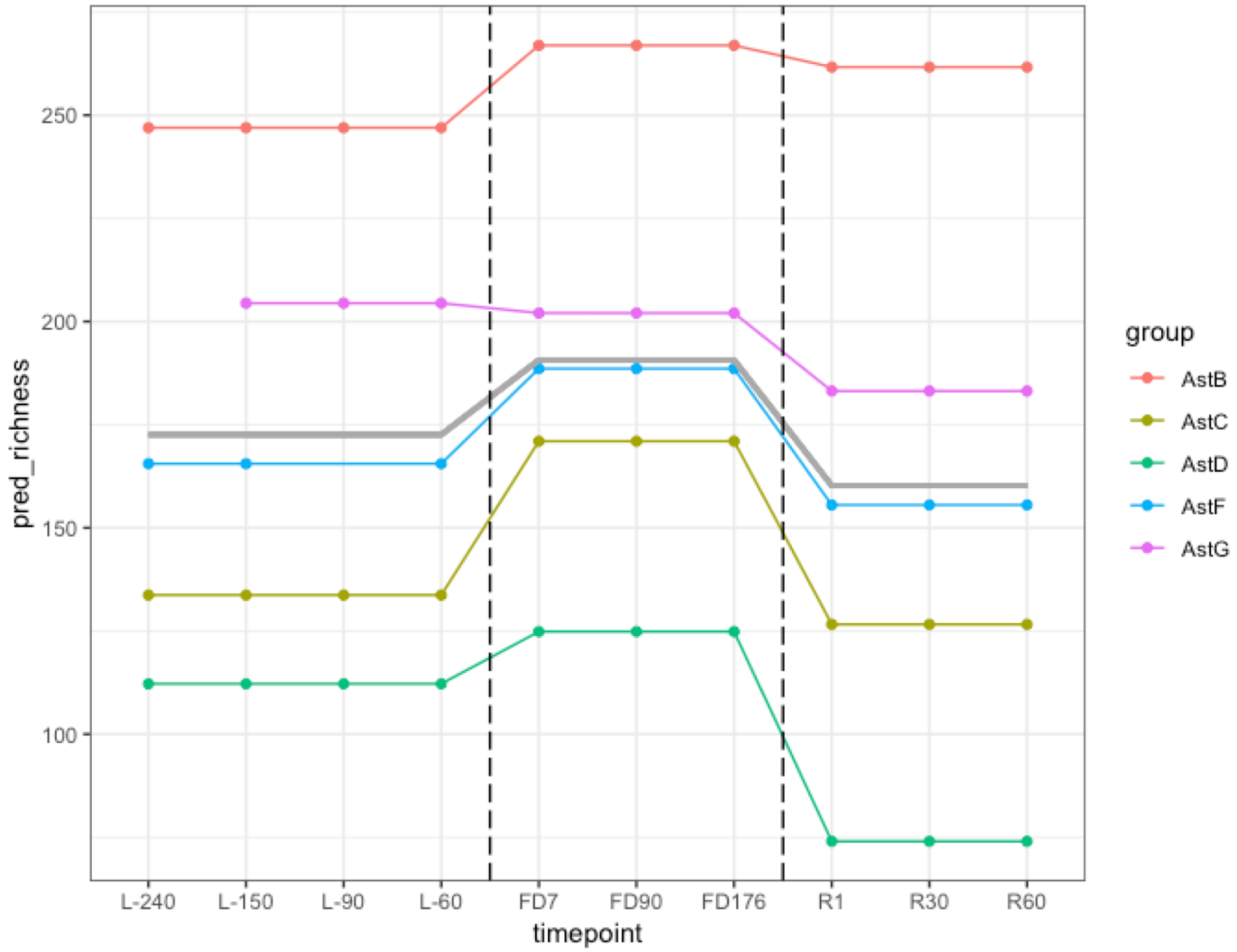
```
> anova(lm.richness)
Analysis of Variance Table

Response: richness
          Df Sum Sq Mean Sq F value Pr(>F)
flight     2   7171  3585.6  0.9696  0.387
Residuals 45 166411  3698.0
```

Mixed-effects Model 2: Random Intercept and Random Slope

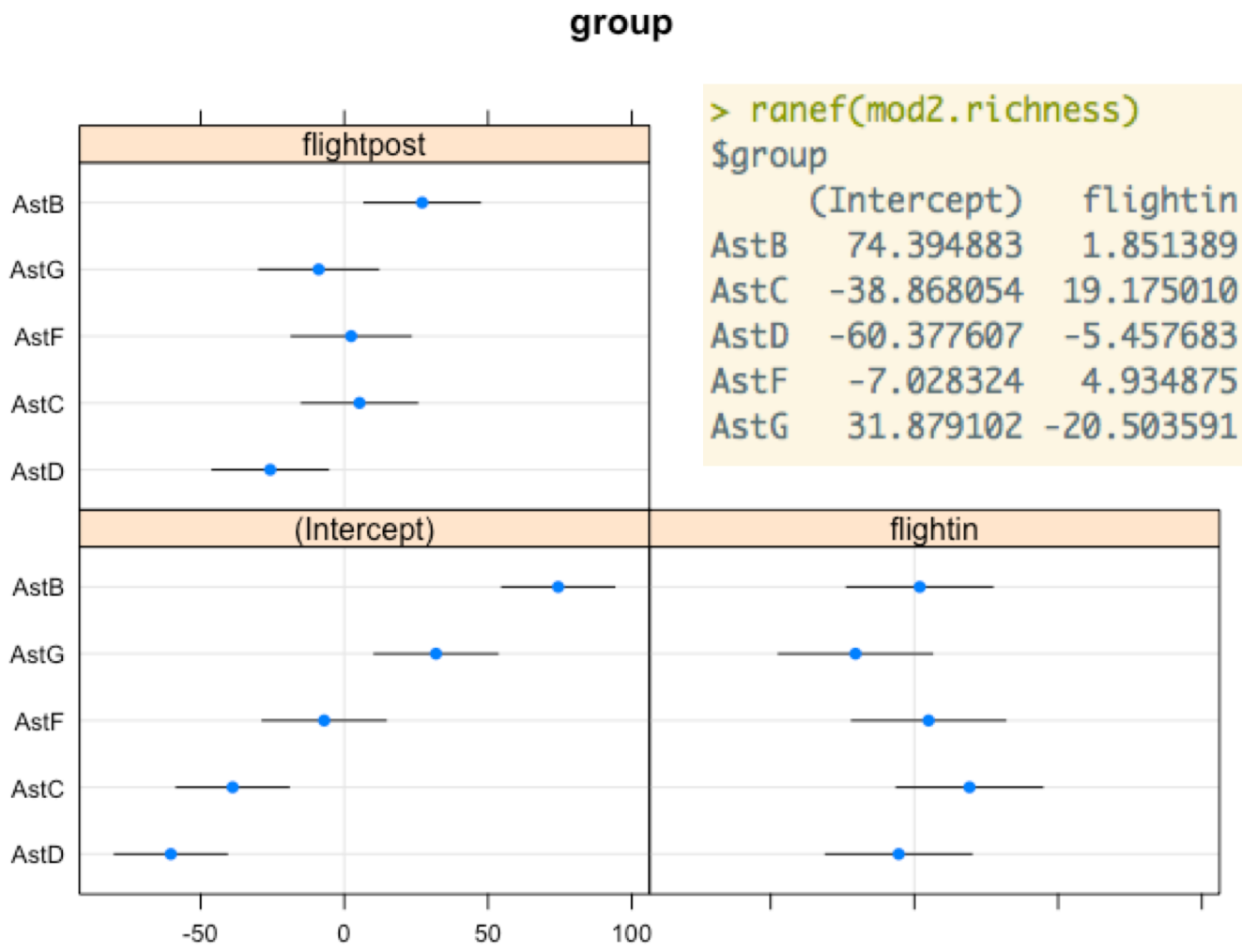
Model specification

```
> mod2.richness <- lmer(richness ~ flight + (flight|group), data = dat)
> summary(mod2.richness)
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: richness ~ flight + (flight | group)
   Data: dat

REML criterion at convergence: 442.6

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.1671 -0.4112  0.0004  0.5225  2.8316

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 group    (Intercept) 3047.2   55.20
          flightin     386.8   19.67    -0.32
          flightpost   486.2   22.05     0.58  0.59
 Residual              592.8   24.35
Number of obs: 48, groups:  group, 5

Fixed effects:                                          p-value
            Estimate Std. Error      df t value Pr(>|t|)
(Intercept)  172.574     25.353   4.031   6.807  0.00237 **
flightin      18.109     12.257   4.088   1.477  0.21213
flightpost   -12.372     13.043   4.651  -0.949  0.38947
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Fitted Values



Prediction of Random Effects




```
> ranef(mod2.richness)
$group
     (Intercept)   flightin flightpost
AstB   74.394883   1.851389  27.070173
AstC  -38.868054  19.175010   5.269282
AstD  -60.377607  -5.457683 -25.781467
AstF   -7.028324   4.934875   2.367180
AstG   31.879102 -20.503591  -8.925168
```

Model Selection

- Two mixed-effects models

```
> anova(mod.richness, mod2.richness)
refitting model(s) with ML (instead of REML)
Data: dat
Models:
mod.richness: richness ~ flight + (1 | group)
mod2.richness: richness ~ flight + (flight | group)
              Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
mod.richness   5 478.18 487.54 -234.09   468.18
mod2.richness 10 483.57 502.29 -231.79   463.57 4.6109      5     0.4652
```

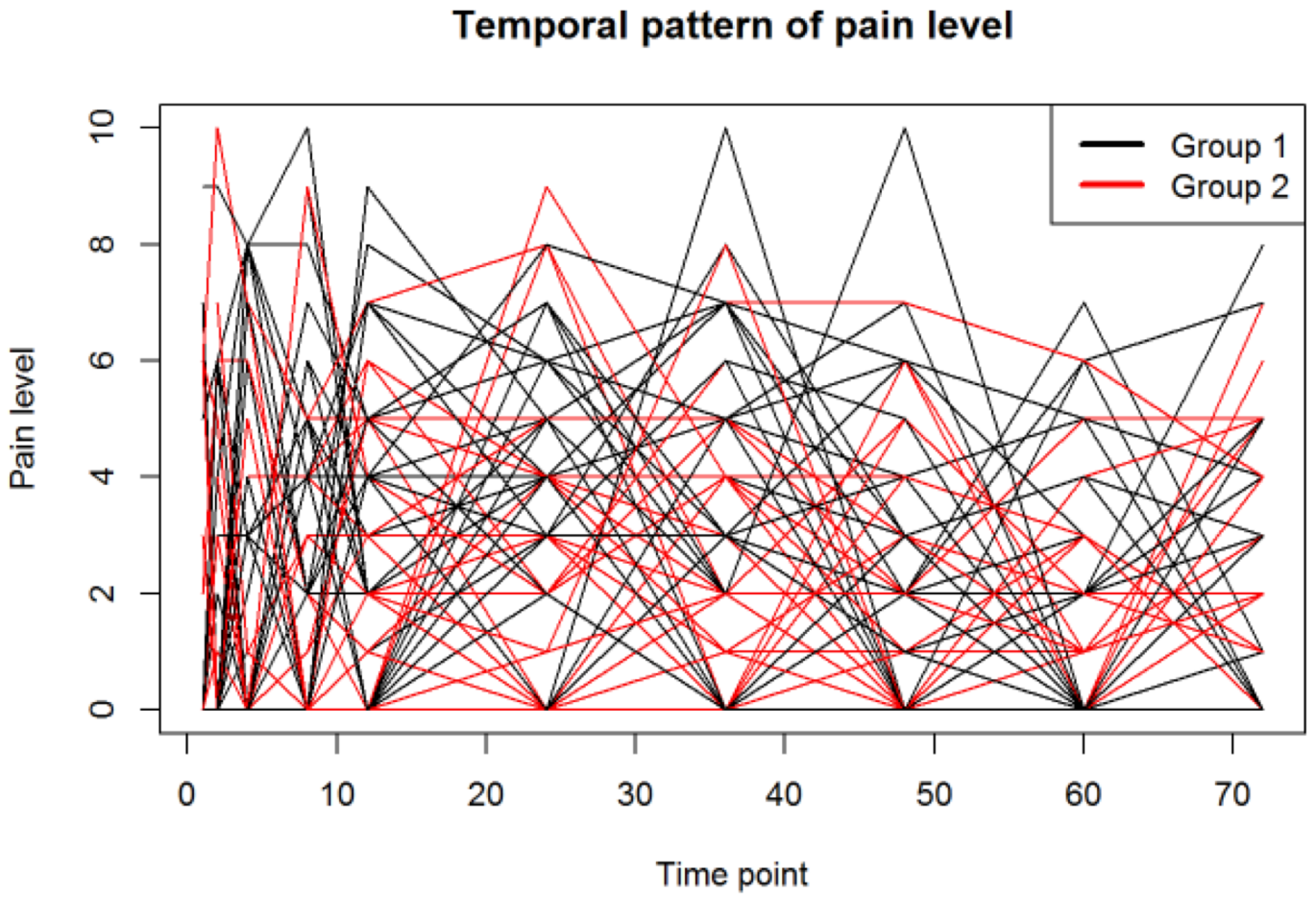
- Mixed-effects model and linear model

```
> anova(mod.richness, lm.richness)
refitting model(s) with ML (instead of REML)
Data: dat
Models:
lm.richness: richness ~ flight
mod.richness: richness ~ flight + (1 | group)
             Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
lm.richness   4 535.47 542.95 -263.73   527.47
mod.richness  5 478.18 487.54 -234.09   468.18 59.282      1  1.366e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Example 2: Longitudinal Study

Pain Level Longitudinal Data

- Hypothesis: group 1 (control) and group 2 (treatment) have different pain levels



Group Comparison

- At each time point: t-test or Wilcoxon rank-sum test

```
##                    t.test wilcox.test
## time_point_1  0.53912276   0.9850068
## time_point_2  0.63350036   0.6654815
## time_point_4  0.08237666   0.1311354
## time_point_8  0.06939804   0.1189577
## time_point_12 0.34713925   0.5330165
## time_point_24 0.15873530   0.1376273
## time_point_36 0.25048237   0.3537464
## time_point_48 0.22456618   0.2304420
## time_point_60 0.73178624   0.7004888
## time_point_72 0.83185413   0.5410899
```

- Use mixed-effects model to build an overarching model for all time points
  - Fixed effects: group, time point
  - Random effects: random intercept for each patient

Mixed-effects Model with Random Intercept

```
## overall: LMER
mod.H2 <- lmer(pain_level ~ time_pt + group + (1|record_id), data=myResponses)
summary(mod.H2)
```

```
## Linear mixed model fit by REML t-tests use Satterthwaite approximations
##   to degrees of freedom [lmerMod]
## Formula: pain_level ~ time_pt + group + (1 | record_id)
##    Data: myResponses
##
## REML criterion at convergence: 3556
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.7927 -0.7254 -0.3633  0.6663  3.1195
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  record_id (Intercept) 0.7635   0.8738
##  Residual              5.3233   2.3072
## Number of obs: 771, groups:  record_id, 79
##
## Fixed effects:
##              Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)  2.060692   0.201171 123.800000  10.244   <2e-16 ***
## time_pt      0.004012   0.003388 693.900000   1.184   0.2368
## group2      -0.539218   0.257699  78.000000  -2.092   0.0397 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Some Comments

Generalized Linear Mixed Model (GLMM)

- Generalized linear model "+" mixed-effects model
- It is an extension to the generalized linear model (GLM) in which the linear predictor contains random effects in addition to the usual fixed effects
- GLM part to account for non-normally distributed data
- Random effects part to account for additional source of variation in correlated data

- To summarize the **regression topic**