Longitudinal Data: Repeated Measures

Mixed Models 1

Alan Hubbard

Oct 25, 2018

Read in data

```
dat1 = read_csv("deeks_ex1.csv")
```

```
## Parsed with column specification:
## cols(
##   id = col_integer(),
##   cd4 = col_integer(),
##   binage = col_integer()
## )
```

```
tbl_df(dat1)
```

```
## # A tibble: 594 x 3
##       id   cd4 binage
##    <int> <int>  <int>
##  1     1    45      0
##  2     1   119      0
##  3     2   196      1
##  4     2   369      1
##  5     4    29      0
##  6     4   137      0
##  7     5    84      1
##  8     5    93      1
##  9     7   246      0
## 10     7   439      0
## # ... with 584 more rows
```

Binary time-independent variable (balanced data)

Random Intercept Model

```r
# random intercept
lme_1 <- lmer(cd4 ~ binage + (1 | id), data = dat1)
summary(lme_1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: cd4 ~ binage + (1 | id)
##    Data: dat1
##
## REML criterion at convergence: 7445.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.7471 -0.3915 -0.1013  0.3812  3.3980
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id       (Intercept) 24883    157.74
##  Residual              5095     71.38
## Number of obs: 594, groups:  id, 297
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   225.90      13.39  16.871
## binage         24.24      19.23   1.261
##
## Correlation of Fixed Effects:
##        (Intr)
## binage -0.696
```

post-estimation stats

```
library(sjstats)
# ICC - var(b0i)/(var(b0i)+var(eij))
icc(lme_1)
```

```
##
## Linear mixed model
##
## Family : gaussian (identity)
## Formula: cd4 ~ binage + (1 | id)
##
##   ICC (id): 0.8300
```

```
AIC(lme_1)
```

```
## [1] 7453.81
```

```
# Now do by grabbing relevant objects
rand_vars1 <- re_var(lme_1)
rand_vars1
```

```
##       Within-group-variance: 5095.062
##      Between-group-variance: 24882.595 (id)
```

```
rand_vars1[2]/(rand_vars1[1] + rand_vars1[2])
```

```
## id_tau.00
##  0.830038
```

Random intercept and slope model

```
# random slope - try it, it won't work (not identifiable,
# why) lme_1b <- lmer(cd4 ~ (binage|id), data=dat1)
```

Binary time-dependent variable (balanced data)

Read in data

```
dat2 = read_csv("deeks_ex2.csv")
```

```
## Parsed with column specification:
## cols(
##   id = col_integer(),
##   cd4 = col_integer(),
##   medvl = col_integer()
## )
```

```
tbl_df(dat2)
```

```
## # A tibble: 142 x 3
##       id   cd4 medvl
##    <int> <int> <int>
##  1    16   449     0
##  2    16   226     1
##  3    18   294     0
##  4    18   138     1
##  5    21   132     1
##  6    21   132     0
##  7    26   324     1
##  8    26   500     0
##  9    30   216     1
## 10    30   254     0
## # ... with 132 more rows
```

Random Intercept Model

```r
# random intercept
lme_2 <- lmer(cd4 ~ medvl + (1 | id), data = dat2)
summary(lme_2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: cd4 ~ medvl + (1 | id)
##    Data: dat2
##
## REML criterion at convergence: 1808.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.0055 -0.4830 -0.0936  0.4303  3.2651
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id       (Intercept) 23820    154.34
##  Residual              8960     94.66
## Number of obs: 142, groups:  id, 71
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   310.66      21.49  14.458
## medvl         -48.28      15.89  -3.039
##
## Correlation of Fixed Effects:
##       (Intr)
## medvl -0.370
```

post-estimation stats

```
# ICC - var(b0i)/(var(b0i)+var(eij))
icc(lme_2)
```

```
##
## Linear mixed model
##
## Family : gaussian (identity)
## Formula: cd4 ~ medvl + (1 | id)
##
##   ICC (id): 0.7267
```

```
AIC(lme_2)
```

```
## [1] 1816.93
```

```
# Now do by grabbing relevant objects
rand_vars2 <- re_var(lme_2)
rand_vars2
```

```
##       Within-group-variance: 8960.088
##      Between-group-variance: 23820.260 (id)
```

```
rand_vars2[2]/(rand_vars2[1] + rand_vars2[2])
```

```
## id_tau.00
## 0.7266628
```

Random intercept and slope model

```
# random slope also not identifiable here lme_2b <- lmer(cd4
# ~ (medvl|id), data=dat2)
```

Continuous time-dependent variables (very unbalanced data)

Read in data

```
dat3 = read_csv("deeks_ex3.csv")
```

```
## Parsed with column specification:
## cols(
##   id = col_integer(),
##   cd4 = col_integer(),
##   logvlbase = col_double(),
##   logvlchange = col_double()
## )
```

```
tbl_df(dat3)
```

```
## # A tibble: 11,300 x 4
##       id   cd4 logvlbase logvlchange
##    <int> <int>     <dbl>       <dbl>
##  1     1    45      2.70        0
##  2     1   119      2.70        2.52
##  3     1   113      2.70       NA
##  4     1    74      2.70        2.50
##  5     1    95      2.70       NA
##  6     1   120      2.70       NA
##  7     1   209      2.70       NA
##  8     1   375      2.70       NA
##  9     1   137      2.70        1.17
## 10     1    NA      2.70        2.38
## # ... with 11,290 more rows
```

Random Intercept Model

```r
# random intercept
lme_3 <- lmer(cd4 ~ logvlbase + logvlchange + (1 | id), data = dat3)
summary(lme_3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: cd4 ~ logvlbase + logvlchange + (1 | id)
##    Data: dat3
##
## REML criterion at convergence: 87637.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.7739 -0.5248 -0.0489  0.4548  8.6669
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id       (Intercept) 35011    187.1
##  Residual             11750    108.4
## Number of obs: 7053, groups:  id, 406
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  506.539     34.484  14.689
## logvlbase    -52.051      8.141  -6.393
## logvlchange  -53.463      1.855 -28.815
##
## Correlation of Fixed Effects:
##             (Intr) lgvlbs
## logvlbase   -0.961
## logvlchange -0.146  0.184
```

post-estimation stats

```
# ICC - var(b0i)/(var(b0i)+var(eij))
icc(lme_3)
```

```
##
## Linear mixed model
##
## Family : gaussian (identity)
## Formula: cd4 ~ logvlbase + logvlchange + (1 | id)
##
##   ICC (id): 0.7487
```

```
AIC(lme_3)
```

```
## [1] 87647.77
```

```
# Now do by grabbing relevant objects
rand_vars3 <- re_var(lme_3)
rand_vars3
```

```
##       Within-group-variance: 11749.751
##      Between-group-variance: 35010.755 (id)
```

Random intercept and slope model

```r
# random slope also not identifiable here
lme_3b <- lmer(cd4 ~ logvlbase + logvlchange + (logvlchange |
    id), data = dat3)
summary(lme_3b)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: cd4 ~ logvlbase + logvlchange + (logvlchange | id)
##    Data: dat3
##
## REML criterion at convergence: 87345.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.1119 -0.5016 -0.0558  0.4388  9.0926
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  id       (Intercept) 27039    164.43
##           logvlchange  1703     41.27   -0.59
##  Residual             10979    104.78
## Number of obs: 7053, groups:  id, 406
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  553.885     28.978  19.114
## logvlbase    -66.685      6.955  -9.588
## logvlchange  -54.153      2.982 -18.160
##
## Correlation of Fixed Effects:
##             (Intr) lgvlbs
## logvlbase   -0.956
## logvlchange -0.167  0.080
```

post-estimation stats

```
icc(lme_3b)
```

```
## Caution! ICC for random-slope-intercept models usually not meaningful. See 'Note' in `?icc`.
```

```
##
## Linear mixed model
##
## Family : gaussian (identity)
## Formula: cd4 ~ logvlbase + logvlchange + (logvlchange | id)
##
##   ICC (id): 0.7112
```

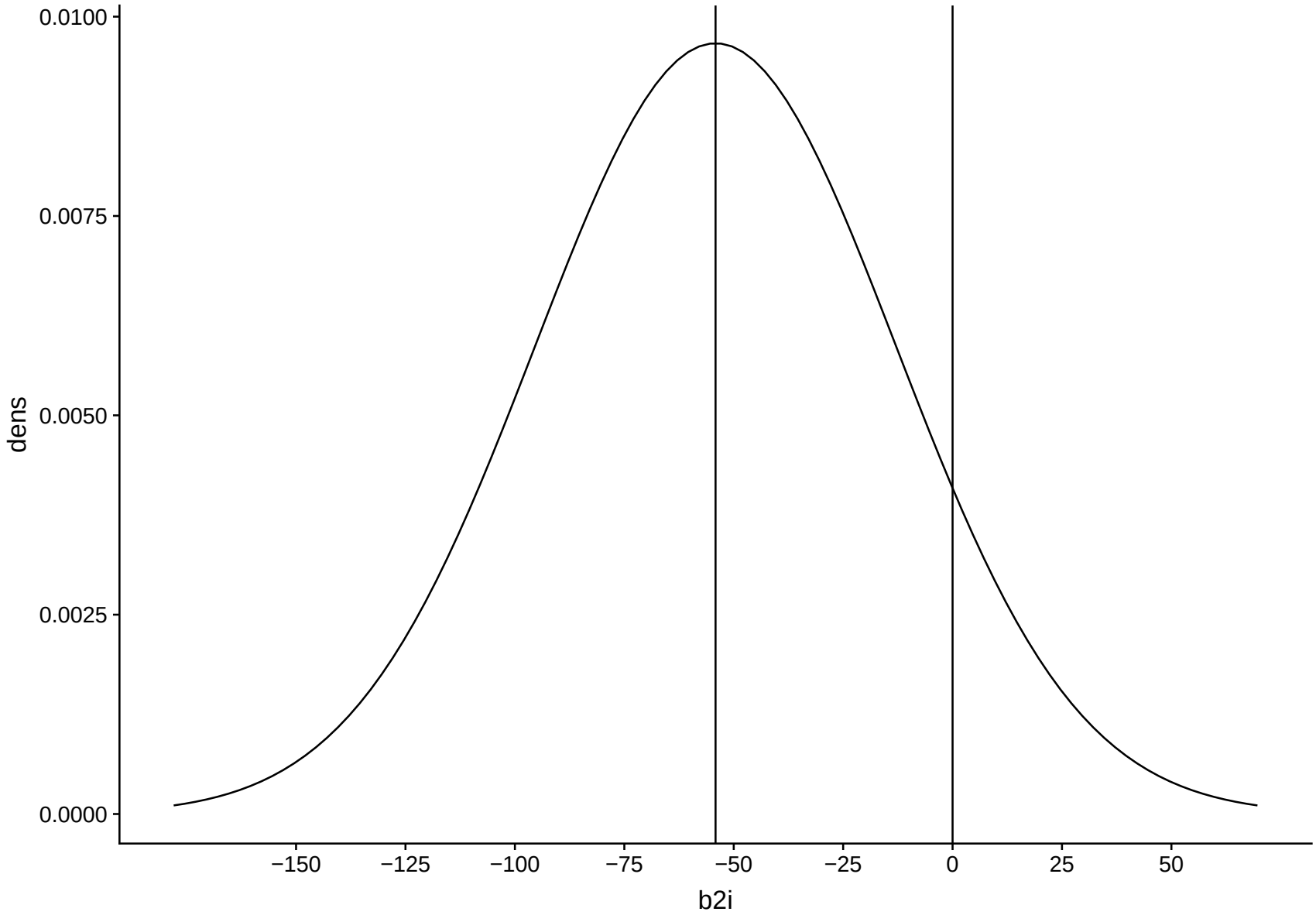
```
AIC(lme_3b)
```

```
## [1] 87359.87
```

```
rand_vars3b <- re_var(lme_3b)
rand_vars3b
```

```
##       Within-group-variance: 10978.987
##      Between-group-variance: 27038.505 (id)
##       Random-slope-variance: 1703.276 (id.logvlchange)
##  Slope-Intercept-covariance: -3976.494 (id.(Intercept))
## Slope-Intercept-correlation:   -0.586 (id)
```

Reporting the estimated distribution of associations from random coefficients model.



Which model does one choose (3 or 3b)?

```
AIC(lme_3, lme_3b)
```

```
##        df      AIC
## lme_3   5 87647.77
## lme_3b  7 87359.87
```

- The random intercept and slope model has lower AIC and thus fits the data relatively better than the random intercept model alone.