Longitudinal Data Analysis

Adam J Sullivan, PhD

04/18/2018

Two Stage Models

Two-Stage (Two-Level) Formulation

- We will proceed with Linear Mixed effects models.
- They are very useful in longitudinal as well as other hierarchical aspects.
- The basic idea of the model is that we assume
  1. **Stage 1**: A straight line (or more generally a "growth" curve) fits the observed responses for each subject.
  2. **Stage 2**: A Regression model relating the mean of the individual intercepts and slopes to the subject specific effects.

Stage 1

- In the first stage we assume that all subjects have their own unique trajectory.
- So for subject i:

$$Y_{ij} = Z_{ij}\beta_i + \varepsilon_{ij}, \qquad j = 1, \ldots, n_i$$

- where β_i is a vector of subject-specific regression parameters, the errors are typically considered independent within a subject.

Stage 1: Subject Specific Effects

- Many times we use a model with subject specific intercepts and slope:

$$Y_{ij} = \beta_{1i} + \beta_{2i} t_{ij} + e_{ij}$$

- So in stage 1 each subject has their own unique regression model.
  - Basically we allow each subject to have their own line.
  - We restrict the covariates in these models to be ones that vary over time.
- Any covariates that do not vary over time or refer to between-subject changes (sex, gender, treatment group, exposure group,…) are not included at this stage.

Stage 2

- In this stage we assume that the β_i's (subject-specific effects) are random and come from some distribution (IE. normal or some other).
- We then model the mean and covariance of the β_i's in the population.

$$\beta_i = A_i\beta + b_i, \text{ where } b_i \sim N(0, G)$$

Stage 2

- Where
  - A_i are the between subject covariates
  - $b_i = \begin{pmatrix} b_{1i} \\ b_{2i} \end{pmatrix}$ are the random effects for individuals
  - $G = \begin{bmatrix} var(b_{1i}) & cov(b_{1i}, b_{2i}) \\ cov(b_{1i}, b_{2i}) & var(b_{2i}) \end{bmatrix}$ is the covariance matrix for the subject specific effects.

Quick Example

- Consider a treatment vs control setting where we have subject specific intercept, β_{1i}, and slope β_{2i}.
- Then we would model the subject specific effects with a group effect:

$$E(\beta_{1i}) = \beta_1 + \beta_2 \mathrm{GROUP}_i$$
$$E(\beta_{2i}) = \beta_3 + \beta_4 \mathrm{GROUP}_i$$

Quick Example

- Where GROUP_i is an indicator variable for treatment.
- Then in this example we would have the following models for means:

Quick Example

- For the control group:

$$E(\beta_{1i}) = \beta_1$$
$$E(\beta_{2i}) = \beta_3$$

Quick Example

- for the treatment group:

$$E(\beta_{1i}) = \beta_1 + \beta_2$$
$$E(\beta_{2i}) = \beta_3 + \beta_4$$

How do we fit these models:

- One approach has been coined as the "NIH Method" since it was popularized by statisticians working at the NIH.
- What they did was:
  1. Fit a regression to the response data for each subject.
  2. Regress the estimates of the individual intercepts and slopes on subject specific covariates.
- This method was very easy to perform because it did not require any special form of regression software.
- This works very well with balanced data.

Mixed Effects Models

Mixed Effects Models

- In contrast what we tend to do now is consider a model that contains the 2 stages but fits everything all at once:

$$
\begin{aligned}
Y_{ij} &= Z_{ij}\beta_i + \varepsilon_{ij} \\
&= Z_{ij}(A_i\beta + b_i) + \varepsilon_{ij} \\
&= Z_{ij}A_i\beta + Z_{ij}b_i + \varepsilon_{ij} \\
&= X_{ij}\beta + Z_{ij}b_i + \varepsilon_{ij}
\end{aligned}
$$

Mixed Effects Models

- We then have:
  - $X_{ij}\beta$ fixed effects (population)
  - $Z_{ij}b_i$ random effects (individual)

An Example

- To illustrate this we consider a study done on orthodontic measurement.
- Investigators at the University of North Carolina Dental School followed the growth of 27 children (16 males, 11 females) from age 8 until age 14.
- Every two years they measured the distance between the pituitary and the pterygomaxillary fissure, two points that are easily identified on x-ray exposures of the side of the head.

An Example

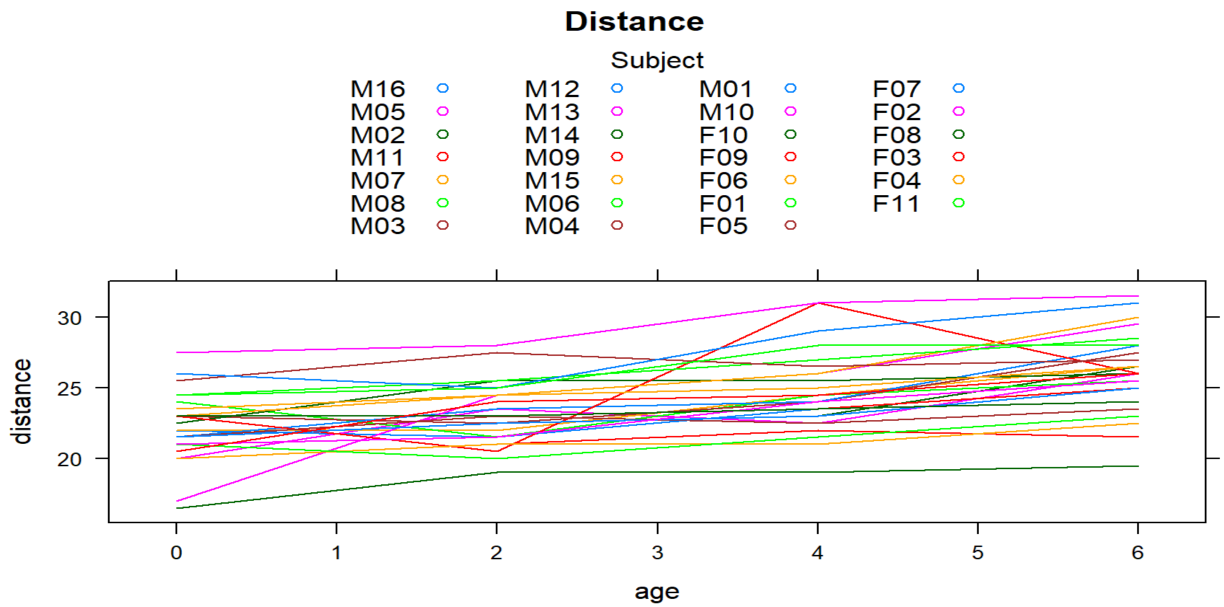
```
library(nlme)
head(Orthodont)
Orthodont$age <- Orthodont$age - 8
```

```
## Grouped Data: distance ~ age | Subject
##   distance age Subject  Sex
## 1     26.0   8     M01 Male
## 2     25.0  10     M01 Male
## 3     29.0  12     M01 Male
## 4     31.0  14     M01 Male
## 5     21.5   8     M02 Male
## 6     22.5  10     M02 Male
```

Example: Another Spaghetti Plot

```
library(lattice)
xyplot( distance ~ age , data= Orthodont, groups=Subject, type='l', auto.key=list(space="top", columns=4
                         title="Subject", cex.title=1), main="Distance")
```

Example: Another Spaghetti Plot



19/47

What do you see?

2 Stage Approach

- Now in the 2 stage approach we first would model the change in distance for each individual.

```
library(nlme)
reg.list <- lmList(distance ~ age, data=Orthodont)
summary(reg.list)
```


```
library(nlme)
reg.list <- lmList(distance ~ age, data=Orthodont)
summary(reg.list)
```

2 Stage Approach

```
Call:
  Model: distance ~ age | Subject
   Data: Orthodont

Coefficients:
   (Intercept)
    Estimate Std. Error t value Pr(>|t|)
M16     21.4        1.1    19.5 4.36e-26
M05     20.4        1.1    18.7 3.32e-25
M02     21.1        1.1    19.2 8.51e-26
M11     22.7        1.1    20.7 2.60e-27
M07     21.4        1.1    19.5 4.36e-26
M08     22.8        1.1    20.8 2.10e-27
M03     22.0        1.1    20.1 1.05e-26
M12     21.2        1.1    19.4 5.44e-26
M13     18.4        1.1    16.8 4.37e-23
M14     23.3        1.1    21.3 6.64e-28
M09     22.2        1.1    20.3 6.79e-27
M15     22.5        1.1    20.5 3.57e-27
M06     24.4        1.1    22.2 7.81e-29
M04     26.1        1.1    23.8 2.59e-30
M01     24.9        1.1    22.7 2.62e-29
M10     27.2        1.1    24.9 3.05e-31
```

2 Stage Approach

```
   age
    Estimate Std. Error t value Pr(>|t|)
M16    0.550      0.293   1.878 6.58e-02
M05    0.850      0.293   2.902 5.36e-03
M02    0.775      0.293   2.646 1.07e-02
M11    0.325      0.293   1.109 2.72e-01
M07    0.800      0.293   2.731 8.51e-03
M08    0.375      0.293   1.280 2.06e-01
M03    0.750      0.293   2.560 1.33e-02
M12    1.000      0.293   3.414 1.22e-03
M13    1.950      0.293   6.657 1.49e-08
M14    0.525      0.293   1.792 7.87e-02
M09    0.975      0.293   3.328 1.58e-03
M15    1.125      0.293   3.840 3.25e-04
M06    0.675      0.293   2.304 2.51e-02
M04    0.175      0.293   0.597 5.53e-01
M01    0.950      0.293   3.243 2.03e-03
M10    0.750      0.293   2.560 1.33e-02
F10    0.450      0.293   1.536 1.30e-01
F09    0.275      0.293   0.939 3.52e-01
F06    0.375      0.293   1.280 2.06e-01
F01    0.375      0.293   1.280 2.06e-01
F05    0.275      0.293   0.939 3.52e-01
```

Abstract Coefficients

- We can then abstract the estimated model coefficients and the variance-covariance matrices:

```
b <- lapply(reg.list, coef)
b
V <- lapply(reg.list, vcov)
V
```

Abstract Coefficients

```
## $M16
## (Intercept)         age
##       21.35        0.55
##
## $M05
## (Intercept)         age
##       20.45        0.85
##
## $M02
## (Intercept)         age
##      21.050       0.775
##
## $M11
## (Intercept)         age
##      22.650       0.325
##
## $M07
## (Intercept)         age
##        21.4         0.8
##
## $M08
## (Intercept)         age
##      22.750       0.375
```

Abstract Coefficients

```
V <- lapply(reg.list, vcov)
V
```

```
## $M16
##             (Intercept)     age
## (Intercept)       0.508 -0.1088
## age              -0.109  0.0363
##
## $M05
##             (Intercept)    age
## (Intercept)       1.417 -0.304
## age              -0.304  0.101
##
## $M02
##             (Intercept)     age
## (Intercept)       0.761 -0.1631
## age              -0.163  0.0544
##
## $M11
##             (Intercept)     age
## (Intercept)      0.2013 -0.0431
## age             -0.0431  0.0144
##
```

Abstract Coefficients

- An indicator variable of the estimate type (alternating intercept and slope) and a subject id variable are also needed, which can be created with:

```
estm <- rep(c("intercept","slope"), length(b))
estm
subj <- rep(names(b), each=2)
subj
```

Abstract Coefficients

```
##  [1] "intercept" "slope"     "intercept" "slope"     "intercept"
##  [6] "slope"     "intercept" "slope"     "intercept" "slope"
## [11] "intercept" "slope"     "intercept" "slope"     "intercept"
## [16] "slope"     "intercept" "slope"     "intercept" "slope"
## [21] "intercept" "slope"     "intercept" "slope"     "intercept"
## [26] "slope"     "intercept" "slope"     "intercept" "slope"
## [31] "intercept" "slope"     "intercept" "slope"     "intercept"
## [36] "slope"     "intercept" "slope"     "intercept" "slope"
## [41] "intercept" "slope"     "intercept" "slope"     "intercept"
## [46] "slope"     "intercept" "slope"     "intercept" "slope"
## [51] "intercept" "slope"     "intercept" "slope"
##  [1] "M16" "M16" "M05" "M05" "M02" "M02" "M11" "M11" "M07" "M07" "M08"
## [12] "M08" "M03" "M03" "M12" "M12" "M13" "M13" "M14" "M14" "M09" "M09"
## [23] "M15" "M15" "M06" "M06" "M04" "M04" "M01" "M01" "M10" "M10" "F10"
## [34] "F10" "F09" "F09" "F06" "F06" "F01" "F01" "F05" "F05" "F07" "F07"
## [45] "F02" "F02" "F08" "F08" "F03" "F03" "F04" "F04" "F11" "F11"
```

Variance Covariance

- Next, we create one long vector with the model coefficients and the corresponding block-diagonal variance-covariance matrix with (the metafor package needs to be loaded for the bldiag() function):

```
library(metafor)
b <- unlist(b)
V <- bldiag(V)
```

Variance Covariance

```
## Error in apply(sapply(mlist, dim), 1, cumsum): dim(X) must have a positive length
```

```
## M16.(Intercept)         M16.age M05.(Intercept)         M05.age
##          21.350           0.550          20.450           0.850
## M02.(Intercept)         M02.age M11.(Intercept)         M11.age
##          21.050           0.775          22.650           0.325
## M07.(Intercept)         M07.age M08.(Intercept)         M08.age
##          21.350           0.800          22.750           0.375
## M03.(Intercept)         M03.age M12.(Intercept)         M12.age
##          22.000           0.750          21.250           1.000
## M13.(Intercept)         M13.age M14.(Intercept)         M14.age
##          18.400           1.950          23.300           0.525
## M09.(Intercept)         M09.age M15.(Intercept)         M15.age
##          22.200           0.975          22.500           1.125
## M06.(Intercept)         M06.age M04.(Intercept)         M04.age
##          24.350           0.675          26.100           0.175
## M01.(Intercept)         M01.age M10.(Intercept)         M10.age
##          24.900           0.950          27.250           0.750
## F10.(Intercept)         F10.age F09.(Intercept)         F09.age
##          17.150           0.450          20.300           0.275
## F06.(Intercept)         F06.age F01.(Intercept)         F01.age
##          20.000           0.375          20.250           0.375
## F05.(Intercept)         F05.age F07.(Intercept)         F07.age
```

Variance Covariance

```
##        [,1]    [,2]   [,3]   [,4]   [,5]    [,6]    [,7]    [,8]   [,9]
##  [1,]  0.508 -0.1088  0.000  0.000  0.000  0.0000  0.0000  0.0000  0.000
##  [2,] -0.109  0.0363  0.000  0.000  0.000  0.0000  0.0000  0.0000  0.000
##  [3,]  0.000  0.0000  1.417 -0.304  0.000  0.0000  0.0000  0.0000  0.000
##  [4,]  0.000  0.0000 -0.304  0.101  0.000  0.0000  0.0000  0.0000  0.000
##  [5,]  0.000  0.0000  0.000  0.000  0.761 -0.1631  0.0000  0.0000  0.000
##  [6,]  0.000  0.0000  0.000  0.000 -0.163  0.0544  0.0000  0.0000  0.000
##  [7,]  0.000  0.0000  0.000  0.000  0.000  0.0000  0.2013 -0.0431  0.000
##  [8,]  0.000  0.0000  0.000  0.000  0.000  0.0000 -0.0431  0.0144  0.000
##  [9,]  0.000  0.0000  0.000  0.000  0.000  0.0000  0.0000  0.0000  0.508
## [10,]  0.000  0.0000  0.000  0.000  0.000  0.0000  0.0000  0.0000 -0.109
## [11,]  0.000  0.0000  0.000  0.000  0.000  0.0000  0.0000  0.0000  0.000
## [12,]  0.000  0.0000  0.000  0.000  0.000  0.0000  0.0000  0.0000  0.000
## [13,]  0.000  0.0000  0.000  0.000  0.000  0.0000  0.0000  0.0000  0.000
## [14,]  0.000  0.0000  0.000  0.000  0.000  0.0000  0.0000  0.0000  0.000
## [15,]  0.000  0.0000  0.000  0.000  0.000  0.0000  0.0000  0.0000  0.000
## [16,]  0.000  0.0000  0.000  0.000  0.000  0.0000  0.0000  0.0000  0.000
## [17,]  0.000  0.0000  0.000  0.000  0.000  0.0000  0.0000  0.0000  0.000
## [18,]  0.000  0.0000  0.000  0.000  0.000  0.0000  0.0000  0.0000  0.000
## [19,]  0.000  0.0000  0.000  0.000  0.000  0.0000  0.0000  0.0000  0.000
## [20,]  0.000  0.0000  0.000  0.000  0.000  0.0000  0.0000  0.0000  0.000
## [21,]  0.000  0.0000  0.000  0.000  0.000  0.0000  0.0000  0.0000  0.000
## [22,]  0.000  0.0000  0.000  0.000  0.000  0.0000  0.0000  0.0000  0.000
```

Final Model

- Finally, we conduct a multivariate meta-analysis with the model coefficients (since we have two correlated coefficients per subject). -The V matrix contains the variances and covariances of the sampling errors.
- We also allow for heterogeneity in the true outcomes (i.e., coefficients) and allow them to be correlated (by using an unstructured variance-covariance matrix for the true outcomes).

Final Model

- The model can be fitted with:

```
res2 <- rma.mv(b ~ estm-1, V, random = ~ estm | subj, struct="UN")
summary(res2)
```

Final Model

```
Multivariate Meta-Analysis Model (k = 54; method: REML)

  logLik  Deviance       AIC       BIC      AICc
-64.4574  128.9148  138.9148  148.6710  140.2192

Variance Components:

outer factor: subj (nlvls = 27)
inner factor: estm (nlvls = 2)

            estim    sqrt  k.lvl  fixed      level
tau^2.1    8.3710  2.8933     27     no  intercept
tau^2.2    0.0478  0.2187     27     no      slope
```

Final Model

```
          rho.intr  rho.slop    intr  slop
intercept        1    0.7394       -    no
slope       0.7394         1      27     -

Test for Residual Heterogeneity:
QE(df = 52) = 1611.6315, p-val < .0001

Test of Moderators (coefficient(s) 1:2):
QM(df = 2) = 3080.0214, p-val < .0001
```

Final Model

```
Model Results:

               estimate      se     zval    pval    ci.lb    ci.ub
estmintercept   26.8868  0.5980  44.9609  <.0001  25.7148  28.0589  ***
estmslope        0.5762  0.0555  10.3868  <.0001   0.4675   0.6850  ***

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

What do we have?

We have:

- We have an estimated average intercept of $b_0 = 26.8868$ (SE=0.5980)
- an estimated average slope of $b_1 = 0.5762$ (SE=0.0555)
- estimated standard deviations of the underlying true intercepts and slopes equal to SD(b_{0i})=2.8933 and SD(b_{1i})=0.2187 , respectively.
- A correlation between the underlying true intercepts and slopes equal to $\hat{\rho} = 0.7394$ (no residual standard deviation is given, since that source of variability is already incorporated into the V matrix).

Mixed Effects Model

Alternative with a Mixed Effects Model

- Alternatively we could have fit this with a mixed model:

```
reg.mix <- lme(distance ~ age, random = ~ age | Subject, data=Orthodont)
summary(reg.mix)
```

Alternative with a Mixed Effects Model

```
Linear mixed-effects model fit by REML
 Data: Orthodont
  AIC BIC logLik
  455 471   -221

Random effects:
 Formula: ~age | Subject
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev Corr
(Intercept) 2.875  (Intr)
age         0.226  0.767
Residual    1.310
```

Alternative with a Mixed Effects Model

```
Fixed effects: distance ~ age
            Value Std.Error DF t-value p-value
(Intercept) 27.32     0.634 80    43.1       0
age          0.66     0.071 80     9.3       0
 Correlation: 
    (Intr)
age 0.762 

Standardized Within-Group Residuals:
     Min       Q1      Med       Q3      Max 
-3.22311 -0.49376  0.00732  0.47215  3.91603 

Number of Observations: 108
Number of Groups: 27
```

What do we see?

- The estimated average distance at age 8 is $b_0 = 27.32$ millimeters (SE=0.634).
- For each year, the distance is estimated to increase on average by $b_1 = 0.66$ millimeters (SE=.071). - However, there is variability in the intercepts and slopes, as reflected by their estimated standard deviations (SD($_{0i}$)=2.875 and SD(b_{1i})=0.226, respectively). Also, intercepts and slopes appear to be somewhat correlated ($\hat{\rho} = 0.767$).
- Finally, residual variability remains (reflecting deviations of the measurements from the subject-specific regression lines), as given by the residual standard deviation of $\hat{\sigma} = 1.310$.

How did this model compare?

- Notice that when we fit this with one model we have smaller standard errors.
- With this approach we are using all of the data at the same time and fitting them together.
- When the model is correctly specified the mixed model approach is preferred.

Adjusting for Sex

- At the same time, it is much easier for us to consider also adjusting for sex.
- This would not be done at stage one but stage 2.
- So in the case of a mixed model we would consider this to be part of the fixed effects but not the random effects:

```
reg.mix2 <- lme(distance ~ age + Sex, random = ~ age | Subject, data=Orthodont)
summary(reg.mix2)
```

Adjusting for Sex

```
Linear mixed-effects model fit by REML
 Data: Orthodont
  AIC BIC logLik
  449 468   -218

Random effects:
 Formula: ~age | Subject
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev Corr
(Intercept) 2.330  (Intr)
age         0.226  0.636
Residual    1.310
```

Adjusting for Sex

```
Fixed effects: distance ~ age + Sex
            Value Std.Error DF t-value p-value
(Intercept) 28.20     0.626 80    45.1   0.000
age          0.66     0.071 80     9.3   0.000
SexFemale   -2.15     0.757 25    -2.8   0.009
 Correlation:
          (Intr) age
age        0.635
SexFemale -0.493  0.000

Standardized Within-Group Residuals:
    Min      Q1     Med      Q3     Max
-3.0814 -0.4568  0.0155  0.4470  3.8944

Number of Observations: 108
Number of Groups: 27
```

What can we see?

- We can see that there does not appear to be a large change in the outcomes by adding sex even though it was significant.
- What we can see that that for Females at the mean age of 8, there is on average a 2.15 mm smaller distance than that of Males who are the same age.