Mixed effects models

Wheels within wheels

Samuel Robinson, Ph.D.

December 10, 2020

Motivation

- What are mixed effects models?
    - Scary math (matrix algebra)
    - Variance partitioning
    - Fixed effects vs. Random effects
- Working with random effects
    - What should be a random effect?
    - Model validation
- Exercise

What are mixed effects models?

Many different names:

1. Mixed effects models
2. Random effects models
3. Hierarchical models
4. Empirical/Bayesian hierarchical models
5. Latent variable models
6. Split-plot models[1]

I like the term *heirarchical models*, as this is the closest to what I will teach you

[1] Earlier form of variance partitioning

Scary math

Unfortunately, we need a review of matrix algebra in order to explain this:

- This is a matrix:

$$A = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$$

- This is a vector

$$b = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$$

- Multiplying them looks like this:

$$A \times b = Ab = 1 \times \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + 2 \times \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} + 3 \times \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix} = \begin{bmatrix} 30 \\ 36 \\ 42 \end{bmatrix}$$

Why do we call them "linear models"?

Involves a *linear mapping* of coefficients onto a model matrix

Coefficients:

$$\beta = \begin{bmatrix} 0.1 & 1.8 & -0.03 \end{bmatrix}$$

Model matrix:

$$X = \begin{bmatrix} 1 & 1 & 10 \\ 1 & 1 & 12 \\ 1 & 0 & 9 \\ \vdots & \vdots & \vdots \end{bmatrix}$$

Multiplying them looks like:

$$\hat{y} = X\beta = \begin{bmatrix} 1.60 \\ 1.54 \\ -0.17 \\ \vdots \end{bmatrix}$$

This is exactly what R does to fit models:

```r
head(dat)
```

```
##           y         x site
## 1 -2.192241 -4.248450    c
## 2  5.735165  5.766103    n
## 3  3.834614 -1.820462    o
## 4 10.403239  7.660348    h
## 5 15.411355  8.809346    a
## 6 -9.085650 -9.088870    g
```

```r
m1 <- lm(y~x,data=dat) #Use variable x to predict y
summary(m1)
```

```
##
## Call:
## lm(formula = y ~ x, data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -25.2451  -4.4342   0.5114   6.2092  18.1923
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.0753     0.6530   1.647    0.102
## x             0.7394     0.1141   6.482 1.27e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.996 on 148 degrees of freedom
## Multiple R-squared:  0.2211, Adjusted R-squared:  0.2159
## F-statistic: 42.02 on 1 and 148 DF,  p-value: 1.267e-09
```

This is exactly what R does to fit models (cont.):

```
head(model.matrix(m1))
```

```
##   (Intercept)         x
## 1           1 -4.248450
## 2           1  5.766103
## 3           1 -1.820462
## 4           1  7.660348
## 5           1  8.809346
## 6           1 -9.088870
```

```
coef(m1)
```

```
## (Intercept)           x
##   1.0753139   0.7393775
```

```
pred2 <- model.matrix(m1) %*% coef(m1) #predicted = matrix * coefs
head(data.frame(pred1=predict(m1),pred2)) #same thing!
```

```
##        pred1      pred2
## 1 -2.0658941 -2.0658941
## 2  5.3386403  5.3386403
## 3 -0.2706944 -0.2706944
## 4  6.7392027  6.7392027
## 5  7.5887456  7.5887456
## 6 -5.6447919 -5.6447919
```

Groups are coded by “dummy variables” (0s and 1s)

```r
m2 <- lm(y~site,data=dat) #Use variable site to predict y
head(model.matrix(m2)) #0s and 1s used to identify groups
```

```
##   (Intercept) siteb sitec sited sitee sitef siteg siteh sitei sitej sitek sitel
## 1           1     0     1     0     0     0     0     0     0     0     0     0
## 2           1     0     0     0     0     0     0     0     0     0     0     0
## 3           1     0     0     0     0     0     0     0     0     0     0     0
## 4           1     0     0     0     0     0     0     1     0     0     0     0
## 5           1     0     0     0     0     0     0     0     0     0     0     0
## 6           1     0     0     0     0     0     1     0     0     0     0     0
##   sitem siten siteo
## 1     0     0     0
## 2     0     1     0
## 3     0     0     1
## 4     0     0     0
## 5     0     0     0
## 6     0     0     0
```

```r
coef(m2) #This uses the 1st site as the "control" group
```

```
## (Intercept)       siteb       sitec       sited       sitee       sitef
##   0.1493889  -0.3476999  -3.1293722   4.7334445  -2.5721233  14.7029246
##       siteg       siteh       sitei       sitej       sitek       sitel
##   1.7083447   2.3612282  -7.8406290  -6.4931198   4.7413122  -3.9443779
##       sitem       siten       siteo
##   5.8387513   0.8629803   1.9148871
```

Structure of LMs... now with matrices!

- All linear models take the form:

$$\hat{y} = X\beta = b_0 1 + b_1 x_1 ... + b_i x_i$$
$$y \sim Normal(\hat{y}, \sigma)$$

- y is a vector of data you want to predict
- $\hat{y}$ is a vector of *predicted values* for y
- $X = \{1, x_1 ...\}$ is a matrix of *predictors* for y
- $\beta = \{b_0, b_1, ...\}$ is a vector of *coefficients*
- $y \sim Normal(\hat{y}, \sigma)$ means:
  - "y follows a Normal distribution with mean $\hat{y}$ and SD σ"

Fixed effects vs. Random effects

Say that X is a model matrix coding for 10 sites[2], and y is something we're interested in predicting

$$\hat{y} = b_0 + X\beta$$
$$y \sim Normal(\hat{y}, \sigma)$$

- Site coefficients (β) are unrelated to each other
- σ is the SD of *residuals*
- Site is a **fixed effect**

$$\hat{y} = b_0 + X\zeta$$
$$y \sim Normal(\hat{y}, \sigma)$$
$$\zeta \sim Normal(0, \sigma_{site})$$

- Site coefficients (β) are related to each other via a *Normal* distribution
- σ is the SD of *residuals*, σ_{site} is the SD of *sites*
- Site is a **random effect**

[2] Intercept is a separate variable

Mixed effects = fixed + random effects

A mixed effects model has both **fixed** and **random** effects

$$\hat{y} = X\beta + U\zeta$$
$$y \sim Normal(\hat{y}, \sigma)$$
$$\zeta \sim Normal(0, \sigma_{site})$$

- X = fixed effects matrix (e.g. intercept, temperature)
- β = fixed effects coefficients
- U = random effects matrix (e.g. sites)
- ζ = random effects coefficients
- σ, σ_{site} = variance terms

Mixed effect model example

Let's go back to our earlier example:

- We're interested in predicting y using x (fixed effects)
- Data was collected at a number of *sites*, which may affect y "somehow"
- Effect of each site is normally distributed

```
head(dat)
```

```
##           y         x site
## 1 -2.192241 -4.248450    c
## 2  5.735165  5.766103    n
## 3  3.834614 -1.820462    o
## 4 10.403239  7.660348    h
## 5 15.411355  8.809346    a
## 6 -9.085650 -9.088870    g
```

```
library(lme4) #Mixed effects library
mm1 <- lmer(y ~ x + (1|site),data=dat) #site is fit as "random intercepts"
```

Mixed effect model example

```r
summary(mm1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: y ~ x + (1 | site)
##    Data: dat
##
## REML criterion at convergence: 1003.2
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.14013 -0.55056  0.09106  0.53825  2.38279
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  site     (Intercept) 23.72    4.870
##  Residual             39.38    6.276
## Number of obs: 150, groups:  site, 15
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  0.99485    1.37791   0.722
## x            0.71548    0.09409   7.604
##
## Correlation of Fixed Effects:
##   (Intr)
## x -0.006
```

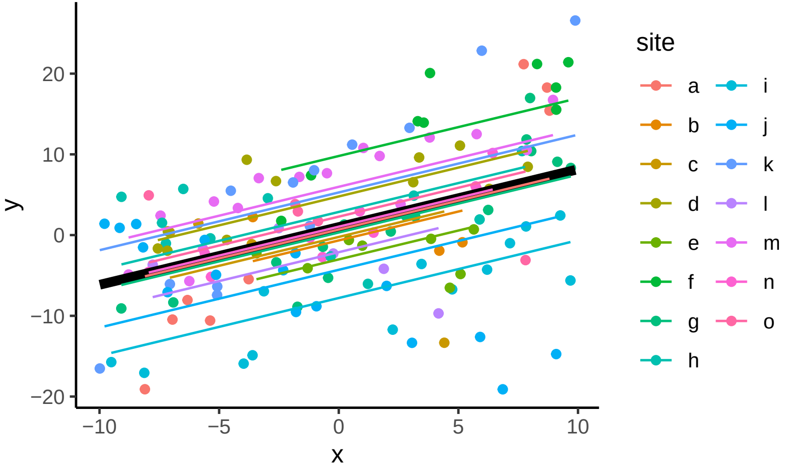
Results from `lmer` model:

- Random effects:
  - *residual* and *site* variance (σ, σ_{site})
- Fixed effects:
  - Intercept and slope estimates (β)
  - No d.f. and p-value [3]

[3] ‘lme4‘ author doesn't think they can be calculated. I largely agree

Mixed effect model results

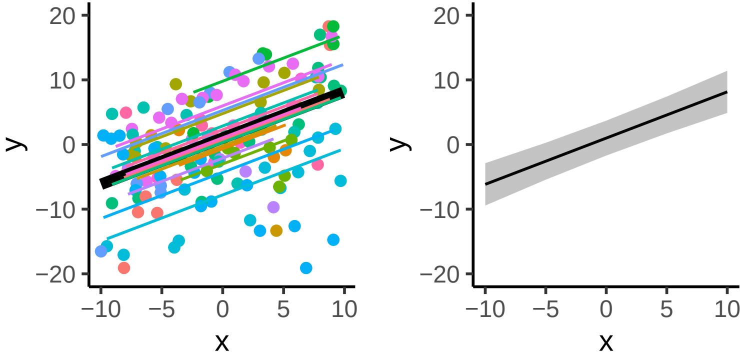
- In a *random intercepts* model, the regression line of x on y is allowed to move up or down around the main regression line for each site
- These changes in intercepts are *normally distributed*



Mixed effect model results (cont.)

For plotting, we want a partial effects plot that marginalizes across sites (i.e. “What does the trend look like at the average site?”)

- `ggpredict` works well for this. If you want partial residuals, you’ll have to add them in yourself using `predict` and `residual`



Random slope + intercept

Suppose that y wasn't just higher or lower at each site, but that the effect of x on y was higher or lower at each site

$$
\begin{aligned}
\hat{y} &= X\beta + U\zeta_{int} + x_i U\zeta_{slope} \\
y &\sim Normal(\hat{y}, \sigma) \\
\zeta_{int} &\sim Normal(0, \sigma_{int}) \\
\zeta_{slope} &\sim Normal(0, \sigma_{slope})
\end{aligned}
$$

- X = fixed effects matrix (e.g. intercept, temperature)
- x_i = individual fixed effect (e.g. temperature)
- β = fixed effects coefficients
- U = random effects matrix (e.g. sites)
- $\zeta_{int}, \zeta_{slope}$ = random intercept and slope coefficients
- $\sigma, \sigma_{int}, \sigma_{slope}$ = variance terms

Random slope and intercept example:

```r
#Intercept varies with site, and slope of x can
#   also vary with site (both hierarchical)
mm2 <- lmer(y ~ x + (x|site),data=dat)
summary(mm2)
```

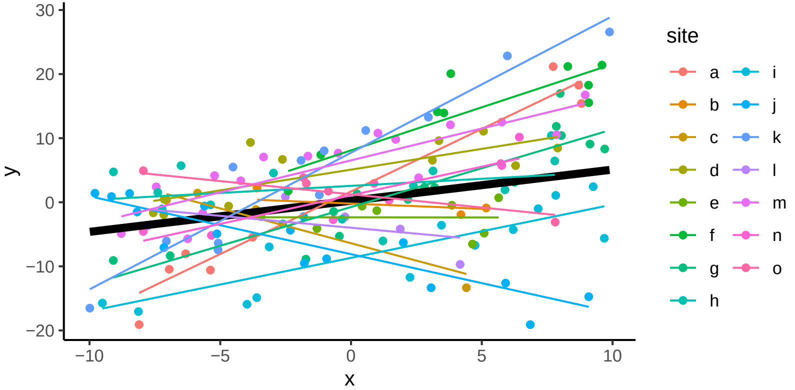
```
## Linear mixed model fit by REML ['lmerMod']
## Formula: y ~ x + (x | site)
##    Data: dat
##
## REML criterion at convergence: 868.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.7003 -0.6290  0.0972  0.5622  2.4476
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  site     (Intercept) 30.695   5.540
##           x            1.012   1.006    0.61
##  Residual             10.781   3.283
## Number of obs: 150, groups:  site, 15
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   0.2376     1.4673   0.162
## x             0.4858     0.2676   1.815
##
## Correlation of Fixed Effects:
##   (Intr)
## x 0.576
```

Results from `lmer` model:

- Random effects:
  - *residual*, *slope*, and *site* variance (σ, σ_{int}, σ_{slope})
  - Correlation b/w intercept and slope = 0.61
    - Sites with higher intercept *also* have a higher slope
- Fixed effects:
  - Intercept and slope estimates

Mixed effect model results

- Regression line of *x* on *y* is allowed to move up or down around the main regression line for each site (random intercepts)
- Slope of regression line can be more or less steep for each site (random slopes)
- Changes in intercepts and slopes are *normally distributed*, and in this example, are correlated with each other



Why do we need to do any of this?

"My supervisor told me to just use site as a fixed effect. Why can't I do that?"

- You can do it this way, but you may encounter the following problems:
  - You lose the *partial pooling* that occurs in mixed effects models = Worse estimates of site effects!
  - You lose 1 d.f. for each site = Type II error $\uparrow$ = You may not find the fixed effect of interest, even if it's there!
  - Sites with low sample sizes may cause your models to break
  - People[4]who have read statistics books published after 1980 may ask questions
- However, if you have a low number of sites (1-10), fixed effects may work better
  - Hard to estimate σ_{site} if number of sites is low
  - Partial pooling doesn't really help if no other sites are available to "borrow strength" from
  - Easier to interpret (p-values, ANOVA, etc.)

[4] e.g. me