

Mixed effects models

Wheels within wheels

Samuel Robinson, Ph.D.

Oct 6, 2023

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 models¹

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

¹Earlier 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:

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

This is exactly what R does to fit models:

```
head(dat)
```

```
##           y  y2           x site
## 1  1.5101095  0 -4.248450    g
## 2  3.7190900  1  5.766103    j
## 3 -4.3737644  1 -1.820462    f
## 4 30.1459331 109  7.660348    n
## 5  0.2777422  1  8.809346    o
## 6 -3.6978175  2 -9.088870    p
```

```
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
## -18.2574  -4.1262   0.0296   3.1854  25.2780
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.7306     0.5772   4.731 4.92e-06 ***
## x              0.7213     0.1020   7.074 4.60e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.3 on 158 degrees of freedom
## Multiple R-squared:  0.2405, Adjusted R-squared:  0.2357
## F-statistic: 50.04 on 1 and 158 DF,  p-value: 4.6e-11
```

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
## 2.7305689  0.7212867
```

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

```
##      pred1      pred2
## 1 -0.3337812 -0.3337812
## 2  6.8895819  6.8895819
## 3  1.4174942  1.4174942
## 4  8.2558758  8.2558758
## 5  9.0846325  9.0846325
## 6 -3.8251119 -3.8251119
```

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

```
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      0      0      0      0      1      0      0      0      0  
## 2          1      0      0      0      0      0      0      0      0      1      0  
## 3          1      0      0      0      0      1      0      0      0      0      0  
## 4          1      0      0      0      0      0      0      0      0      0      0  
## 5          1      0      0      0      0      0      0      0      0      0      0  
## 6          1      0      0      0      0      0      0      0      0      0      0  
##      sitem siten siteo sitep  
## 1      0      0      0      0  
## 2      0      0      0      0  
## 3      0      0      0      0  
## 4      0      1      0      0  
## 5      0      0      1      0  
## 6      0      0      0      1
```

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

```
## (Intercept)      siteb      sitec      sited      sitee      sitef  
##  7.192416 -11.998464 -14.632803   1.983649  -7.765354  -4.523079  
##      siteg      siteh      sitei      sitej      sitek      sitel  
## -3.439621  -8.280601  -4.306456  -4.085855  -5.663021  -5.155112  
##      sitem      siten      siteo      sitep  
## -6.226642   8.403599  -8.626661 -10.934182
```


Structure of LMs... now with matrices!

- All linear models take the form:

$$\hat{y} = X\beta = b_0 1 + b_1 x_1 \dots + b_i x_i$$

$$y \sim \text{Normal}(\hat{y}, \sigma)$$

- y is a vector of data you want to predict
- \hat{y} is a vector of *predicted values* for y
- $X = \{1, x_1 \dots\}$ is a matrix of *predictors* for y
- $\beta = \{b_0, b_1, \dots\}$ is a vector of *coefficients*
- $y \sim \text{Normal}(\hat{y}, \sigma)$ 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 sites², and y is something we're interested in predicting

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

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

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

- Site coefficients (ζ) 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**

²Intercept is a separate variable

Mixed effects = fixed + random effects

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

$$\hat{y} = X\beta + U\zeta$$

$$y \sim \text{Normal}(\hat{y}, \sigma)$$

$$\zeta \sim \text{Normal}(0, \sigma_{\text{site}})$$

- X = fixed effects matrix (e.g. intercept, temperature)
- β = fixed effects coefficients
- U = random effects matrix (e.g. sites)
- ζ = random effects coefficients
- $\sigma, \sigma_{\text{site}}$ = 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  y2           x site
## 1  1.5101095  0 -4.248450    g
## 2  3.7190900  1  5.766103    j
## 3 -4.3737644  1 -1.820462    f
## 4 30.1459331 109  7.660348    n
## 5  0.2777422  1  8.809346    o
## 6 -3.6978175  2 -9.088870    p
```

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

Mixed effect model example

```
summary(mm1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: y ~ x + (1 | site)
## Data: dat
##
## REML criterion at convergence: 1040.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.50816 -0.71380 -0.02682  0.69401  3.01951
##
## Random effects:
## Groups Name Variance Std.Dev.
## site (Intercept) 26.78 5.175
## Residual 31.68 5.628
## Number of obs: 160, groups: site, 16
##
## Fixed effects:
## Estimate Std. Error t value
## (Intercept) 2.00331 1.38297 1.449
## x 0.76192 0.08083 9.426
##
## Correlation of Fixed Effects:
## (Intr)
## x -0.002
```

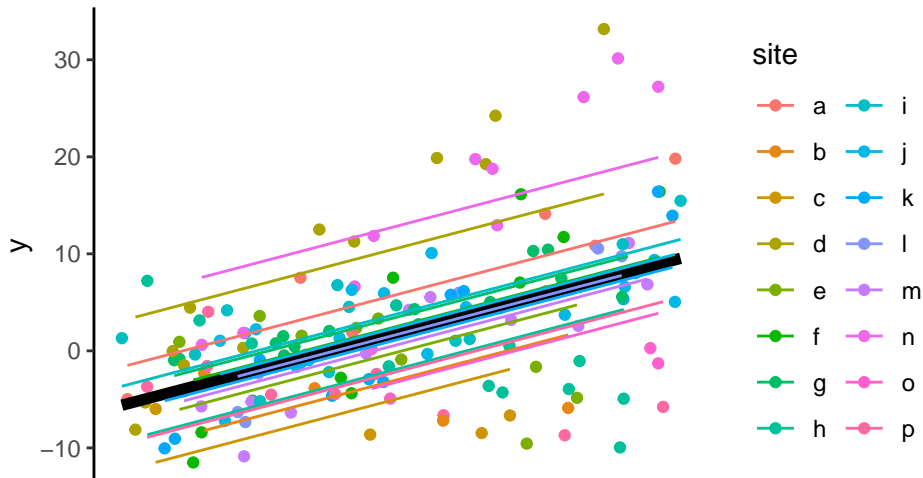
Results from lmer model:

- Random effects:
 - *residual* and *site* variance (σ , σ_{site})
- Fixed effects:
 - Intercept and slope estimates (β)
 - No d.f. and p-value³

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

Mixed effect model results

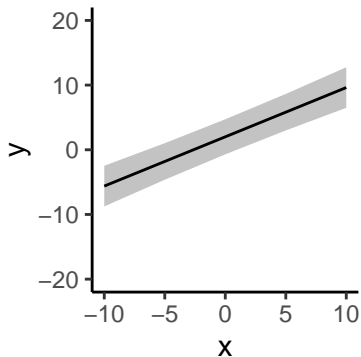
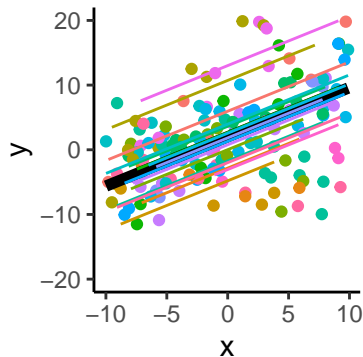
- In a *random intercepts* model, the regression line of x on y is allowed to move up or down around the main regression line for each site
- These changes in intercepts are *normally distributed*



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

$$\hat{y} = X\beta + U\zeta_{int} + x_i U\zeta_{slope}$$

$$y \sim \text{Normal}(\hat{y}, \sigma)$$

$$\zeta_{int} \sim \text{Normal}(0, \sigma_{int})$$

$$\zeta_{slope} \sim \text{Normal}(0, \sigma_{slope})$$

- 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:

```
#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)
```

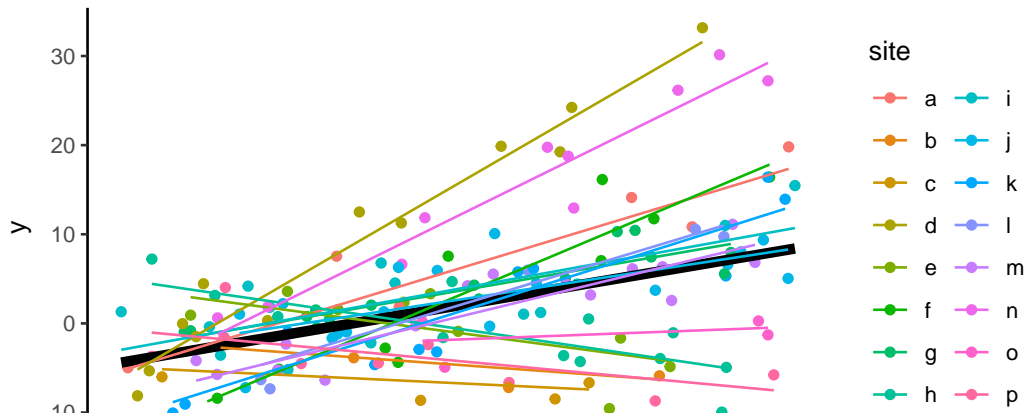
```
## Linear mixed model fit by REML ['lmerMod']  
## Formula: y ~ x + (x | site)  
## Data: dat  
##  
## REML criterion at convergence: 900.6  
##  
## Scaled residuals:  
##      Min       1Q   Median       3Q      Max   
## -2.10500 -0.64857  0.02414  0.61137  2.22996   
##  
## Random effects:  
## Groups   Name      Variance Std.Dev. Corr  
## site     (Intercept) 35.2210  5.9347  
##          x           0.7889  0.8882   0.82  
## Residual                9.3162  3.0522  
## Number of obs: 160, groups: site, 16  
##  
## Fixed effects:  
##              Estimate Std. Error t value  
## (Intercept)   2.0383    1.5091    1.351  
## x             0.6438    0.2275    2.830  
##  
## Correlation of Fixed Effects:  
##      (Intr)  
## x 0.790
```

Results from lmer model:

- Random effects:
 - *residual, slope, and site* variance (σ , σ_{int} , σ_{slope})
 - Correlation b/w intercept and slope = 0.82
 - 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?”

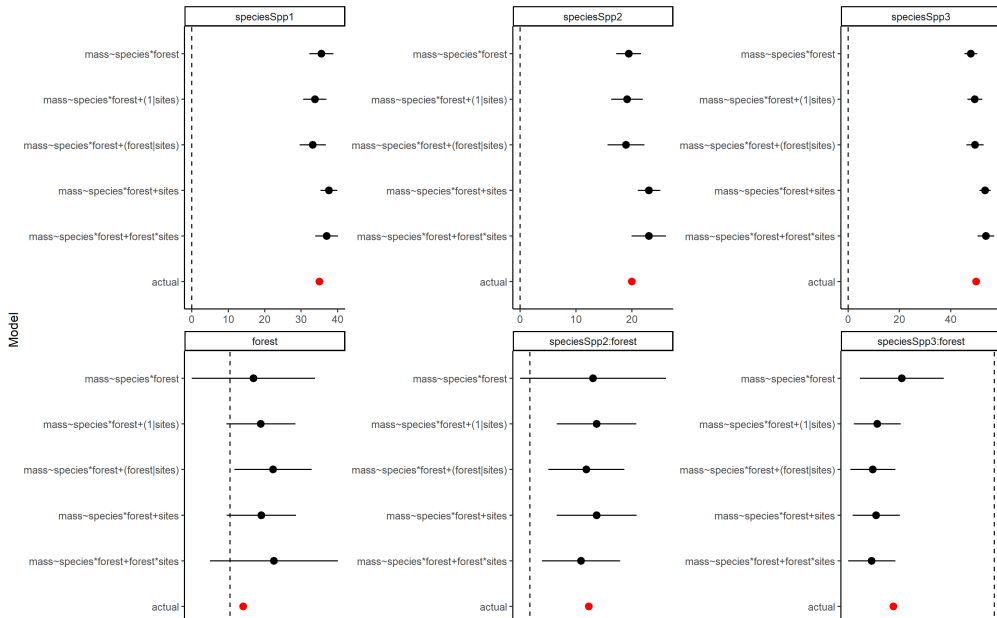
- You can do it this way, but you may encounter the following problems:
 - You lose the *partial pooling* that occurs in mixed effects models = Worse estimates of site effects!
 - You lose 1 d.f. for each site = Type II error \uparrow = You may not find the fixed effect of interest, even if it’s there!
 - Sites with low sample sizes may cause your models to break
 - People⁴who 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.)

⁴e.g. me

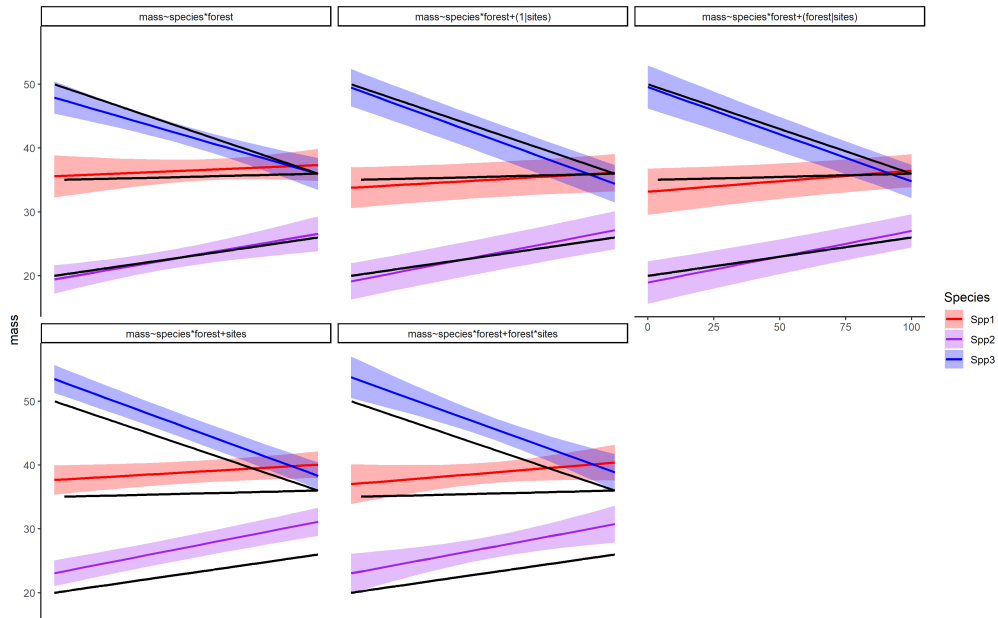
A challenger approaches!

- You're interested in how forest cover influences the mass of three bat species. Maybe some of the species do better in forests?
- You've weighed a bunch of bats across different forest covers (`batMass.csv`). However, these were collected across 15 separate sites. Perhaps some of the variation is just caused by the site?
- Fit a mixed effects model with the fixed effects you're interested in (**forest cover**, **species**), and include **site** as a random effect (*intercept* or *slope*)
- Your supervisor doesn't like hierarchical models, and tells you to just use site as another fixed term in an `lm` model. Do you get different results if you use their approach?

Results:



Results:



Motivation

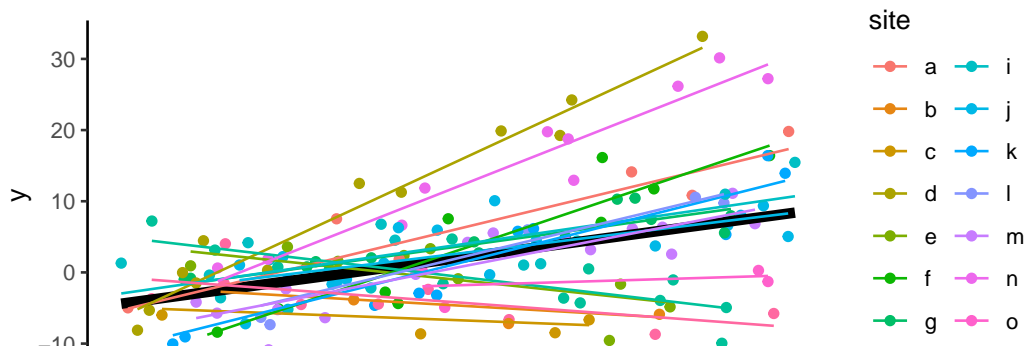
- How do I check if model results are valid?
 - Residual checks
 - Hypothesis testing
- What if my response variable is non-normal?
 - Generalized linear mixed models (GLMMs)
- Sampling over time or space
 - “Continuous” random effects
- Christmas-themed exercise!

Mixed effect model example

Let's go back to our earlier example:

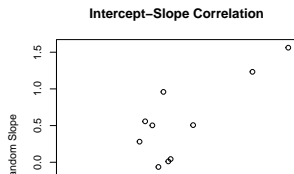
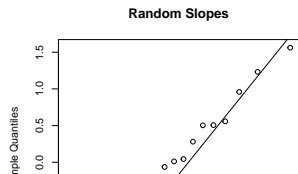
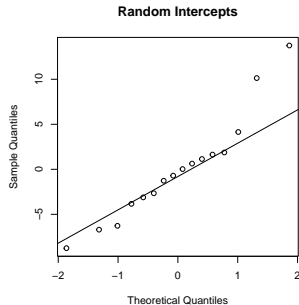
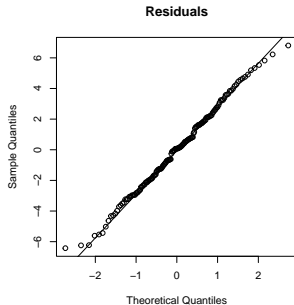
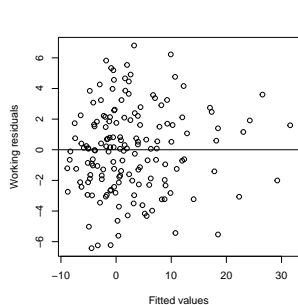
```
lmer(y ~ x + (x|site), data = dat)
```

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



Validation

- Similar to linear models, but we *also* check whether the random intercepts are normally distributed



Hypothesis testing

Is this fixed effect important? (e.g. ANOVA)

- Use likelihood-based test via `drop1` (likelihood ratio test, AIC)
- Be careful to fit model with `REML = FALSE`!

```
lmm1 <- update(lmm1,REML=FALSE) #Refit model using ML rather than REML  
drop1(lmm1,test='Chisq') #x has a strong effect
```

```
## Single term deletions  
##  
## Model:  
## y ~ x + (x | site)  
##          npar      AIC      LRT  Pr(Chi)  
## <none>          913.08  
## x              1 917.85 6.7697 0.009272 **  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hypothesis testing (cont.)

How do I know this effect is different from x ?

- Use Wald Z-test (2-sided p-value from Z-test)

```
lmm1 <- update(lmm1, REML=TRUE) #Reset to REML
meanEst <- fixef(lmm1)[2] #Get mean
seEst <- sqrt(vcov(lmm1)[2,2]) #Get standard error
(1-pnorm(meanEst/seEst,0,1))*2 #p-value from 2-sided Z-test
```

```
##           x
## 0.004659069
```

- glht from library(multcomp) works with lmer models if you are comparing between coefficients (e.g. "Is treatment A different from B and C?")

What if my response variable is non-normal?

- Linear model (LM)

$$\hat{y} = X\beta$$

$$y \sim \text{Normal}(\hat{y}, \sigma)$$

- Linear mixed effects model (LMM)

$$\hat{y} = X\beta + U\zeta$$

$$y \sim \text{Normal}(\hat{y}, \sigma)$$

$$\zeta \sim \text{Normal}(0, \sigma_{\text{site}})$$

- Generalized linear model (GLM)

$$\text{logit}(\hat{\phi}) = X\beta$$

$$y \sim \text{Binomial}(\hat{\phi})$$

- Generalized linear mixed effects model (GLMM)

$$\text{logit}(\hat{\phi}) = X\beta + U\zeta$$

$$y \sim \text{Binomial}(\hat{\phi})$$

$$\zeta \sim \text{Normal}(0, \sigma_{\text{site}})$$

How do I fit GLMMs?

- `glmer` and `glmer.nb` from `library(lme4)` work for Binomial, Poisson, and Negative Binomial data

```
library(lme4)
glmm1 <- glmer.nb(y2~x+(x|site),data=dat) #Negative binomial GLMM
summary(glmm1) #glmer.nb takes a LONG time to run
```

- `glmmTMB` from `library(glmmTMB)` works for those above, *plus* a bunch of others
 - e.g. Zero-inflation, Beta-binomial, Spatial Models

```
library(glmmTMB)
glmm2 <- glmmTMB(y2~x+(x|site),data=dat,family=nbinom2())
summary(glmm2) #Similar results, but quicker
```

Fitting GLMMs - glmer.nb

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##   Family: Negative Binomial(5.1295)  ( log )
## Formula: y2 ~ x + (x | site)
##   Data: dat
##
##           AIC          BIC    logLik deviance df.resid
##        627.8        646.3    -307.9   615.8       154
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.3745 -0.7098 -0.3946  0.5108  2.5367
##
## Random effects:
##   Groups Name            Variance Std.Dev. Corr
##   site  (Intercept)  1.43502   1.1979
##         x              0.02878   0.1697   0.92
## Number of obs: 160, groups:  site, 16
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.32746    0.32166   1.018   0.3087
## x            0.10830    0.04681   2.314   0.0207 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##   (Intr)
## x 0.799
```

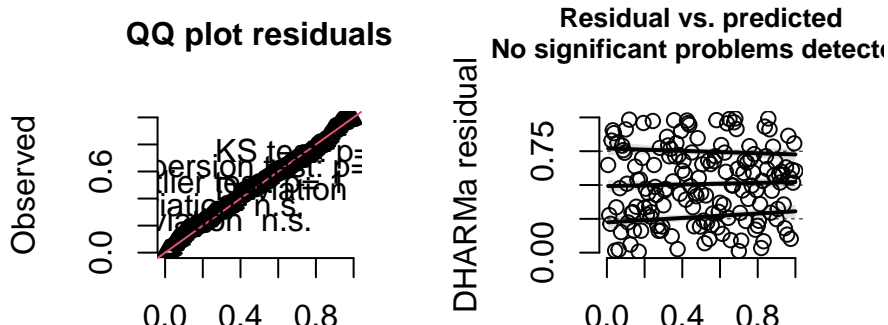
Fitting GLMMs - glmmTMB

```
## Family: nbinom2 ( log )
## Formula:      y2 ~ x + (x | site)
## Data: dat
##
##      AIC      BIC   logLik deviance df.resid
##    627.8    646.2   -307.9   615.8     154
##
## Random effects:
##
## Conditional model:
##   Groups Name      Variance Std.Dev. Corr
##   site  (Intercept) 1.43543  1.1981
##         x           0.02892  0.1701  0.92
## Number of obs: 160, groups: site, 16
##
## Dispersion parameter for nbinom2 family (): 5.12
##
## Conditional model:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) 0.34132    0.32172   1.061  0.2887
## x           0.11026    0.04697   2.348  0.0189 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual checks on glmmTMB models

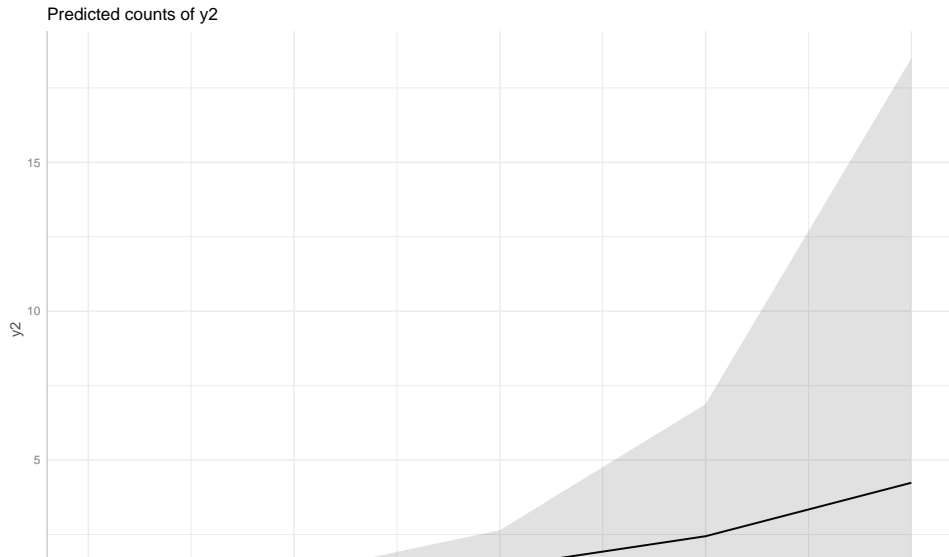
- Extract residuals and make your own plots, or use `simulateResiduals` from `library(DHARMa)` (see [here](#))
- DHARMa also has useful functions for checking overdispersion and zero-inflation (found [here](#))

DHARMa residual



Partial residual plots for glmmTMB models

- `ggpredict()` from `library(ggeffects)` works with `glmmTMB` models



Spatial and Temporal Random Effects

“My data were sampled over time or over a geographic area (or both). Can I just use day or site as a random effect?”

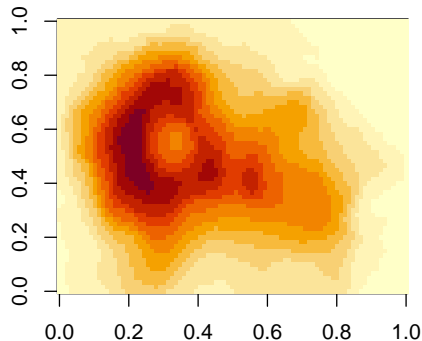
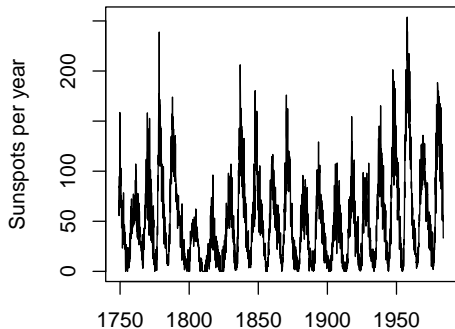
- Short answer: “Yes”
- Long answer: You might be able to do better, because of the **1st Law of Geography**:

“... everything is related to everything else, but near things are more related than distant things.” Waldo Tobler

- If you have spatial or temporal information, this can help R to estimate random effects more accurately
 - Can improve prediction accuracy (smaller p-values)
 - Can give you hints about the underlying causal mechanisms

Temporal or Spatial Data

- Correlation is often present in temporal data or spatial data; causes may be unknown or “uninteresting”
- Usually we are interested in accounting for these patterns, in order to better estimate the “interesting” patterns on top of them



Covariance

- Normal distributions⁵ don't just have a single σ , but a matrix of values
- If our data y are *independent*, then it looks like this:

$$y \sim \text{Normal}(\textcolor{brown}{M}, \textcolor{red}{\Sigma})$$

$$\textcolor{brown}{M} = [\mu_1, \mu_2, \mu_3]$$

$$\textcolor{red}{\Sigma} = \begin{bmatrix} \textcolor{red}{\sigma}^2 & 0 & 0 \\ 0 & \textcolor{red}{\sigma}^2 & 0 \\ 0 & 0 & \textcolor{red}{\sigma}^2 \end{bmatrix}$$

- Zeros mean “ μ_1 , μ_2 , & μ_3 aren't related to each other”
- Diagonal elements = *variance*, off-diagonal = *covariance*

⁵Multivariate Normal

Covariance and Correlation

In real life, things may not be independent from each other. For example:

- $\sigma = 2$ (variance = $\sigma^2 = 4$)
- μ_1 and μ_2 are strongly correlated ($r=0.7$), but μ_3 is not related to anything ($r=0$).
Shown here as a *correlation matrix* (R):

$$R = \begin{bmatrix} 1 & 0.7 & 0 \\ 0.7 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- When multiplied by the variance, this becomes the *covariance matrix* (Σ)

$$\Sigma = \begin{bmatrix} \sigma^2 \times 1 & \sigma^2 \times 0.7 & \sigma^2 \times 0 \\ \sigma^2 \times 0.7 & \sigma^2 \times 1 & \sigma^2 \times 0 \\ \sigma^2 \times 0 & \sigma^2 \times 0 & \sigma^2 \times 1 \end{bmatrix} = \begin{bmatrix} 4 & 2.8 & 0 \\ 2.8 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

Gaussian Process Modelling

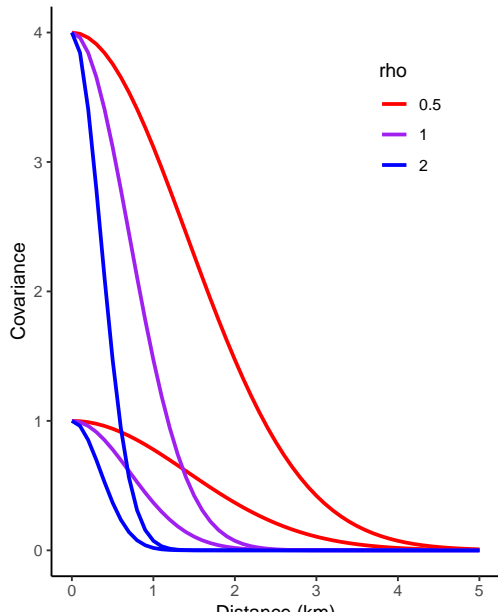
- We can model covariance between things as a function of *distance*, either in time or space
- Squared-exponential is fairly common⁶:

$\Sigma = \text{covariance}$

$\Sigma = \text{variance} \times \text{correlation}$

$\Sigma = \sigma^2 \times e^{-\rho^2 \text{Dist}^2}$

- Instead of finding a single σ value, R now looks for σ (maximum covariance) and ρ (decay with distance)



Spatial random effects

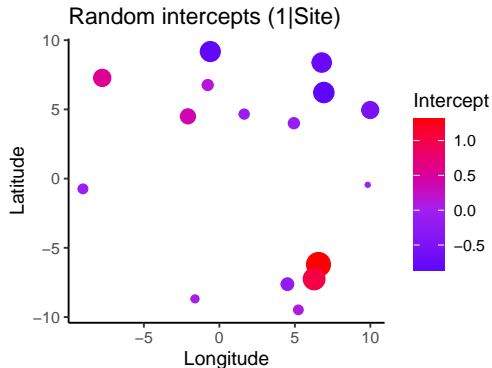
- Say that we collected data at 16 sites, and we're interested in the effect of y on x
- Let's first fit a model with a random intercept for site

#Same syntax as lmer models:

```
lmm2 <- glmmTMB(y~x+(1|site),data=dat2)
```

- If we plot the intercepts for each site, we see that they are clustered:

```
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use `guides()` instead.  
## "none")` instead.
```

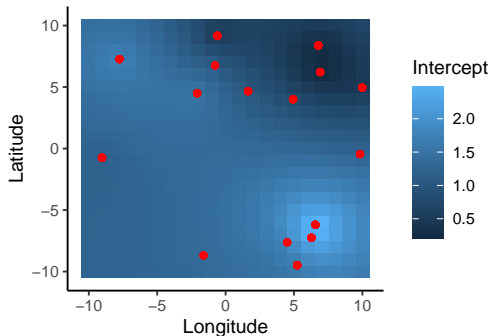


Spatial random effects (cont.)

- Re-fit model with a spatial (exponential) random effect

```
#Coordinates  
dat2$coords <- numFactor(dat2$lon,dat2$lat)  
  
#Group factor (only 1 here)  
dat2$group <- factor(rep(1,nrow(dat2)))  
  
#Fit model with spatial random effect  
lmm3 <- glmmTMB(y~x+exp(coords+0|group),data=dat2)
```

- Clustering effect modeled as a spatial random effect
Spatial random effect



A challenger approaches

- Ho ho ho! Merry Christmas! In order to maximize the number of presents that you get from Santa Claus, you've decided to apply an analytic approach, and have collected data across Alberta on *number of Christmas presents received*
- You've also collected data on things that might influence Saint Nick's generosity (*naughtiness, presence of milk and cookies, chimney width*)
- Fit a GLMM to the present data, one using spatial random intercepts, and one using "regular" random intercepts
- Which type of snack should you leave out for Santa? Which area might you consider moving to??