36-617: Applied Linear Models

mlm residuals

Brian Junker
132E Baker Hall
brian@stat.cmu.edu

Announcements

- HW09 due tonight
- HW10 due next Weds
- Projects:
  - I still owe you grades on Project 01 – working on it!
  - Project 02 out now (HW10, #2: tech appx!)
- Project 02 Schedule:
  - **Wed Nov 17:** Draft Technical Appendix with HW 10.
  - **Wed Nov 24 (or earlier):** Full IDMRAD paper first draft.
  - **Wed Dec 1:** Peer reviews due.
  - **Fri Dec 10 (or earlier):** Full IDMRAD paper final draft!

Plan for rest of semester

- M Nov 8 – intro to MLM’s, continued
- W Nov 10 – residuals for MLM’s
- M Nov 15 – estimation and model selection
- W Nov 17 – shrinkage, crash course on Bayes
- M Nov 22 – catch-up, or multilevel glm’s
- **W Nov 24 – Thanksgiving break!**
- M Nov 29 – ?? Maybe spline smoothing
- W Dec 1 – ?? Maybe spline smoothing

Outline

- The London Schools Data (again!)
  - A nice random-intercepts, random-slopes model
- Residuals in MLM's
  - Marginal residuals
  - Conditional residuals
  - Random effects residuals
- Cholesky Residuals
- The Orthodont Data

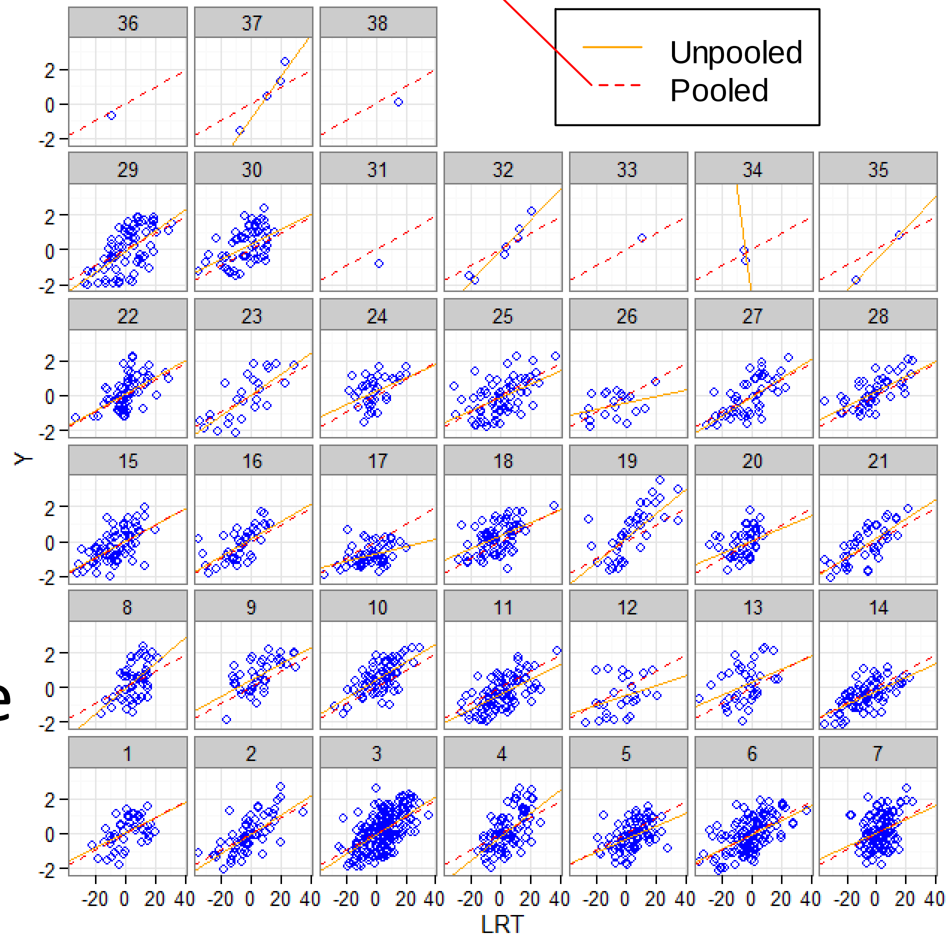
The London Schools Data

- Student (1..1978)
  - Gender (0=Female, 1=Male), per student
  - VR = verbal reasoning level (High/Med/Low)
  - LRT = London Reading test (at beginning of year)
  - Y = end-of-year test

- School (1..38)
  - School.gender (All.Boy, All.Girl, Mixed)
  - School.denom (Other,CofE,RomCath,State)

London Schools Data

- The **dotted line** is the *pooled* regression (ignoring schools)
- The **solid lines** are *unpooled* regressions (separate for each school)
- The solid lines look like a random sample of lines, with “mean” the solid line!



The London Schools Data

- This suggests a model like

$$\begin{aligned}
y_i &= \alpha_{0j[i]} + \alpha_{1j[i]} LRT_i + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2) \\
\alpha_{0j} &= \beta_0 + \eta_{0j}, \quad \eta_{0j} \sim N(0, \tau_0^2) \\
\alpha_{1j} &= \beta_1 + \eta_{1j}, \quad \eta_{1j} \sim N(0, \tau_1^2), \quad j[i] = \text{school for student } i
\end{aligned}$$

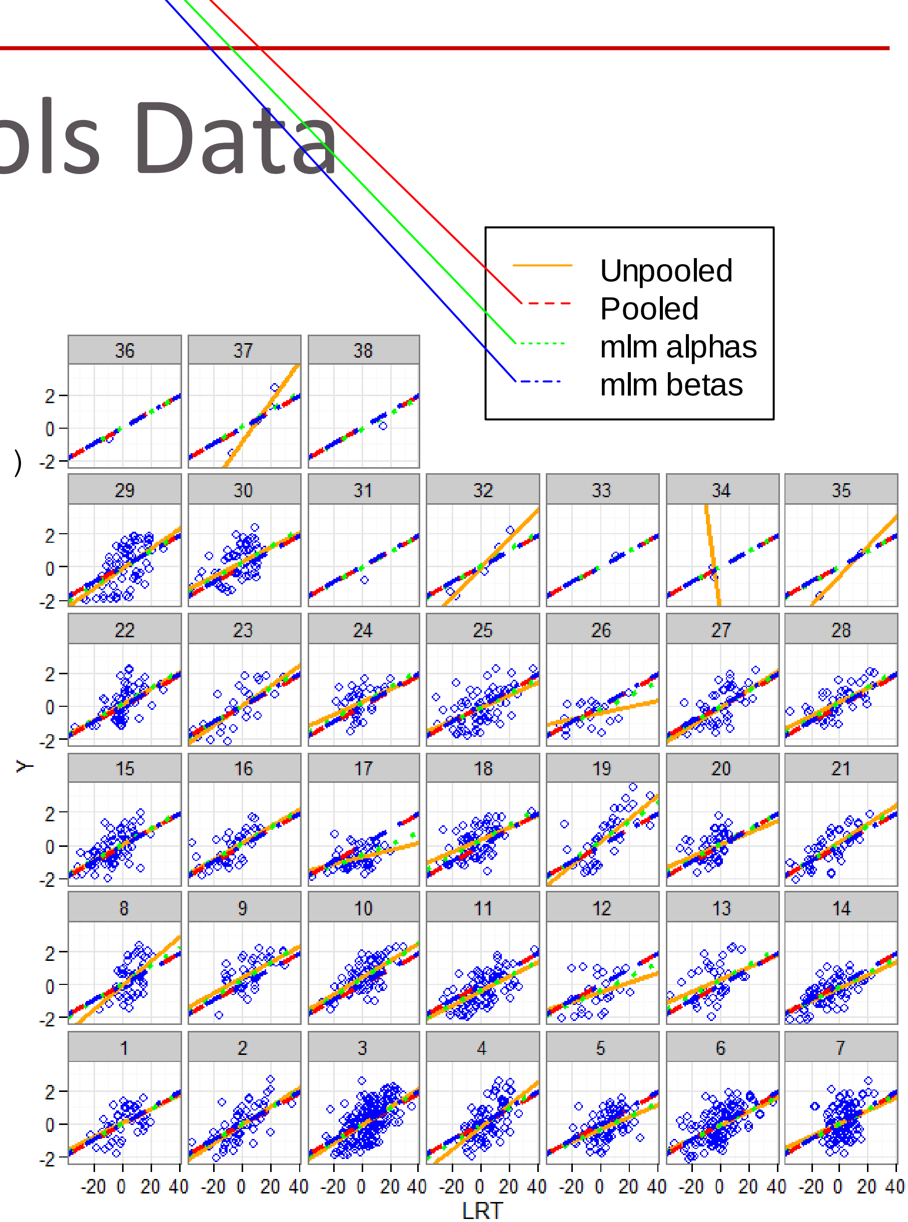
$$y_i = (\beta_0 + \beta_1 LRT_i) + (\eta_{0j[i]} + \eta_{1j[i]} LRT_i) + \epsilon_i$$

- As an R model this would be

  Y ~ 1 + LRT + (1 + LRT|school)

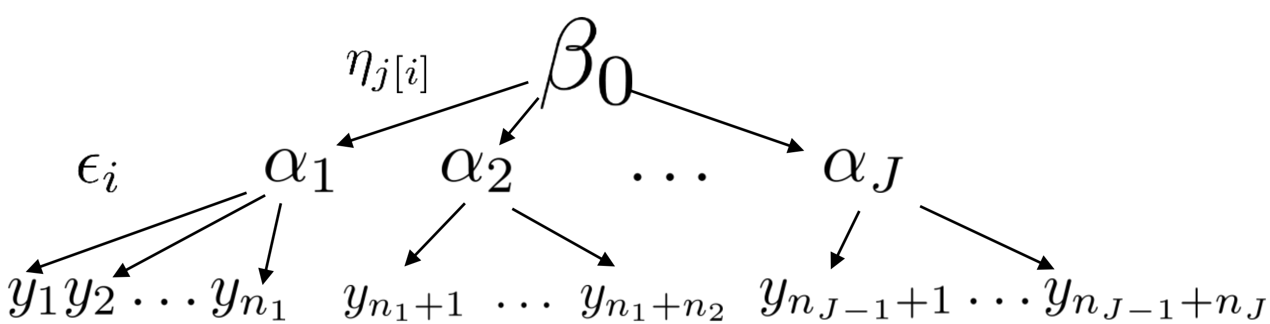
The London Schools Data

```
> display(lmer.1
+ <- lmer(Y ~ 1 + LRT +
+ (1 + LRT|school),data=school.frame))

              coef.est coef.se
(Intercept) 0.01      0.05
LRT         0.05      0.00

Error terms:
 Groups Name  Std.Dev. Corr
 school (Int) 0.23
        LRT   0.01     0.56
 Residual     0.79
---
number of obs: 1978,
groups: school, 38
AIC = 4764.4, DIC = 4722.4
deviance = 4737.4
```

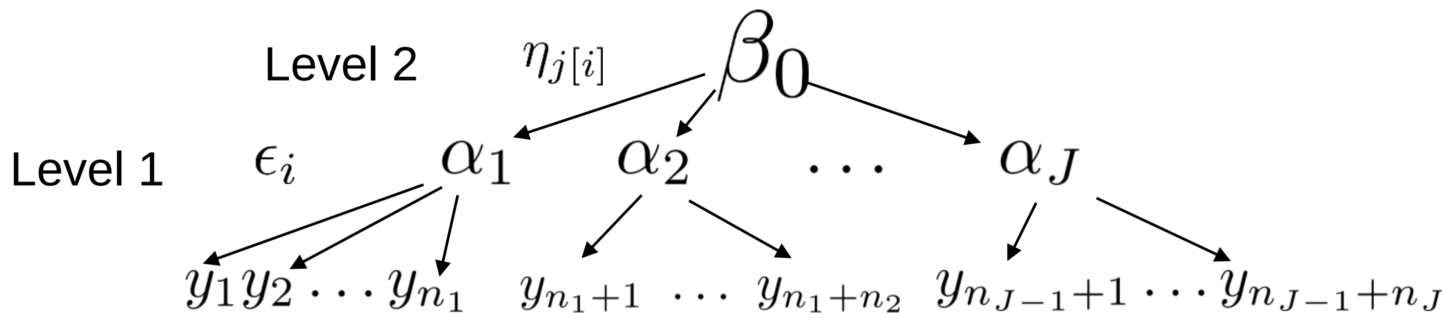


Residuals

- In ordinary linear regression the residuals are easy to think about:
  - $E[y_i] = X_i\beta$
  - $r_i = y_i - E[y_i]$
- Multi-level models pose a couple of challenges



Residuals in Multi-Level Models



- Where are they?
  - Level 1? Level 2? Some combination?
- What are they? The α's are random draws, so does the following make sense?
  - $E[y_{ij}] = \alpha_{0j} + \alpha_{1j}\, LRT_{ij}$??
  - $r_{ij} = y_{ij} - E[y_{ij}]$??

Residuals in Multi-Level Models

- The variance components version of the model

$$y_i = (\beta_0 + \beta_1 LRT_i) + (\eta_{0j[i]} + \eta_{1j[i]} LRT_i) + \epsilon_i$$

could be re-expressed in matrix form as

$$y = X\beta + Z\eta + \epsilon$$

i.e.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & LRT_1 \\ 1 & LRT_2 \\ \vdots & \vdots \\ 1 & LRT_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \left[\begin{array}{cccccc} 1 & LRT_1 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & LRT_{n_1} & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & LRT_{n_1+1} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & LRT_{n_2} & 0 & 0 \\ \hline & & & \ddots & \ddots & \\ \hline 0 & 0 & 0 & 0 & 1 & LRT_{n_{J-1}+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 1 & LRT_{n_J} \end{array}\right] \left[\begin{array}{c} \eta_{01} \\ \eta_{11} \\ \hline \eta_{02} \\ \eta_{12} \\ \hline \vdots \\ \hline \eta_{0J} \\ \eta_{1J} \end{array}\right] + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Laird & Ware (1982, *Biometrics*)

Residuals in Multi-Level Models

- Given the Laird-Ware form $y = X\beta + Z\eta + \epsilon$, can formulate 3 different kinds of residuals:
  - Marginal residuals: $y - X\beta \;\; (\text{``=''} Z\eta + \epsilon)$
  - Conditional residuals: $y - X\beta - Z\eta \;\; (\text{``=''} \epsilon)$
  - Random effects: $y - X\beta - \epsilon \;\; (\text{``=''} Z\eta)$
- In practice, estimate β with $\hat{\beta}$, the MLE, and estimate η with $\eta_{BLUP} \approx E[\eta|\text{the data}]$
  - Although there is only one η per group, there are as many $\hat{\eta}_{BLUP}$'s as there are observations.
- Nobre & Singer (2007, *Biometrical Journal*)

Residuals in Multi-Level Models

- ***Marginal residuals:*** $y - X\beta \quad (\text{“}{=}\text{”}\, Z\eta + \epsilon)$
  - Should be mean 0, but may show grouping structure
  - May not be homoskedastic! *Will be correlated*[1]!
  - Good for checking fixed effects, just like linear regr.
- ***Conditional residuals:*** $y - X\beta - Z\eta \quad (\text{“}{=}\text{”}\, \epsilon)$
  - Should be mean zero with no grouping structure
  - Should be homoskedastic!
  - Good for checking normality of ϵ, outliers
- ***Random effects:*** $y - X\beta - \epsilon \quad (\text{“}{=}\text{”}\, Z\eta)$
  - Will generally not be mean-zero
  - May not be be homoskedastic!
  - Good for checking normality of η, outliers

[1]We'll look at uncorrelated "Cholesky residuals'" in a few slides.

Residuals in the London Schools Data

```
> str(fixef(lmer.1))
> beta0 <- fixef(lmer.1)[1]
> beta1 <- fixef(lmer.1)[2]

> str(ranef(lmer.1))
> eta <- ranef(lmer.1)$school

> attach(school.frame)
> X <- cbind(1,LRT)

> blocks <- lapply(split(X,school),
+   function(x){matrix(x,ncol=2)})
> J <- length(blocks)
> n <- dim(school.frame)[1]

> Z <- matrix(0,nrow=n,ncol=J*2)
```

```
> row <- 1
> for (j in 1:J) {
+   col <- 2*j
+   nj <- dim(blocks[[j]])[1]
+   Z[row:(row+nj-1),c(col-1,col)] <-
+       blocks[[j]]
+   row <- row + nj
+ }

> beta <- rbind(beta0,beta1)
> # so beta is a column vector

> eta <- c(t(eta))
> # so eta is a column vector

> resid.marg <- Y - X%*%beta

> resid.cond <- Y - X%*%beta - Z%*%eta

> resid.reff <- Z%*%eta
```

The file “residual-functions.r” provides these residuals automatically.

Residuals in the London Schools Data

- Marginal residuals

$$y - X\beta \quad (\text{``=''}\, Z\eta + \epsilon)$$

look pretty good…

- Conditional residuals

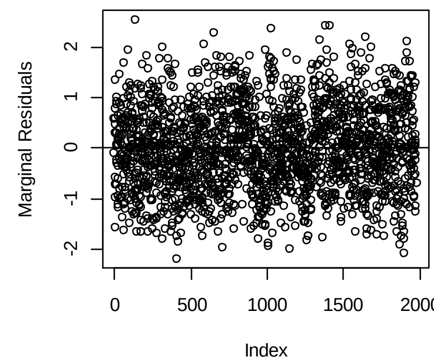
$$y - X\beta - Z\eta \quad (\text{``=''}\, \epsilon)$$

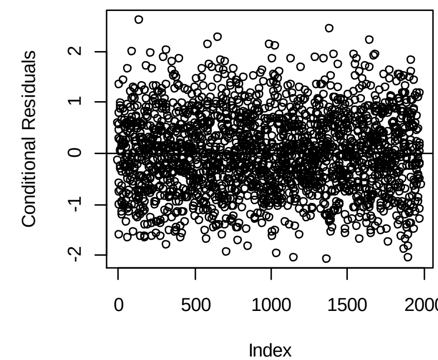
look pretty good

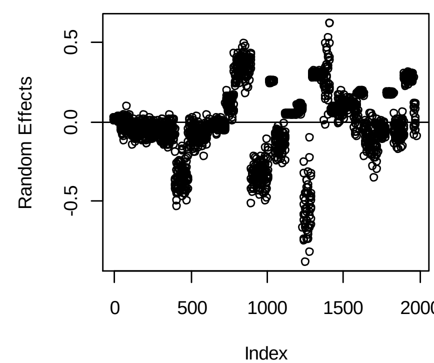
- Rand Effect residuals

$$y - X\beta - \epsilon \quad (\text{``=''}\, Z\eta)$$

look weird…







Residuals in the London Schools Data

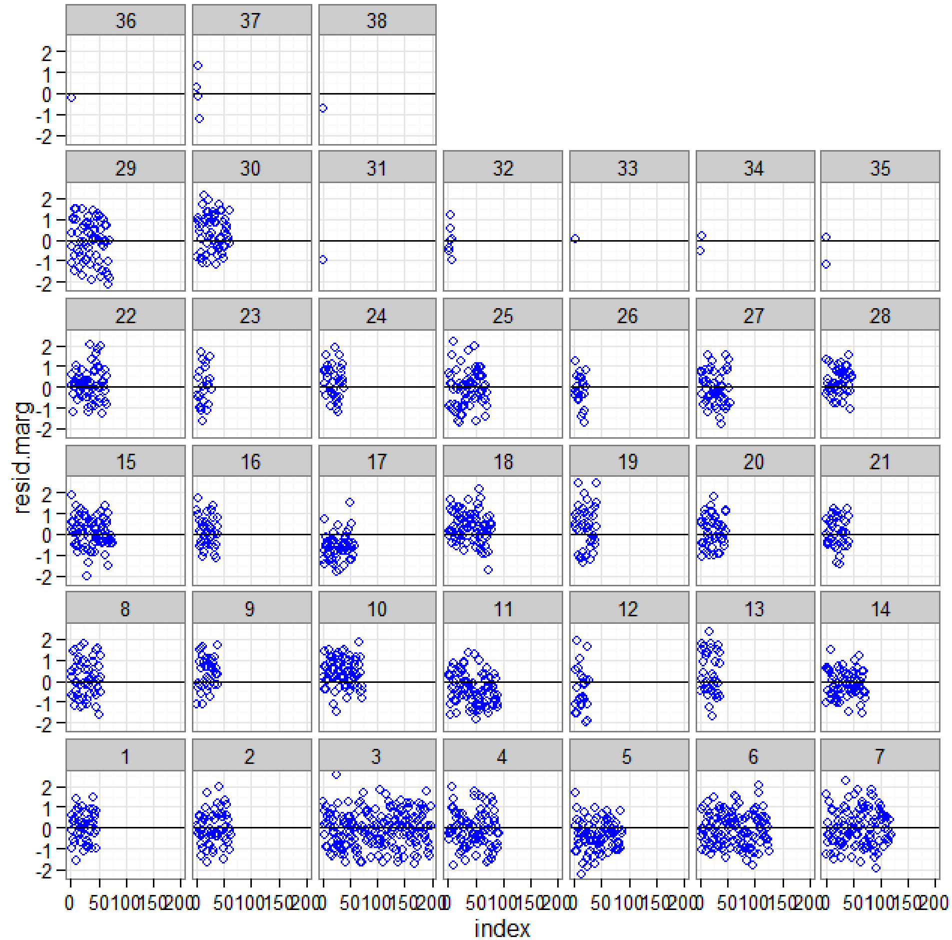
- Marginal residuals

$$y - X\beta \quad (\text{``}{=}\text{''}\, Z\eta + \epsilon)$$

plotted by school

- No noticeable patterns

- Nice set of residuals



Residuals in the London Schools Data

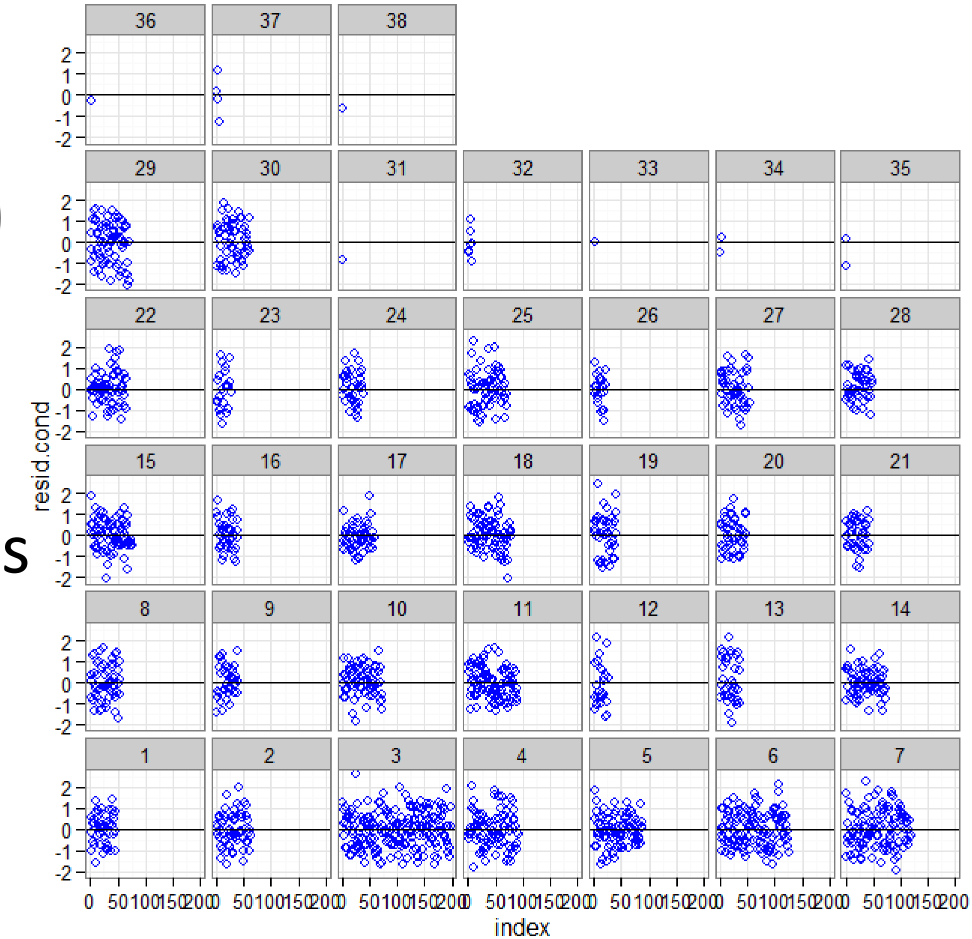
- Conditional residuals

  $$y - X\beta - Z\eta \quad (\text{“}{=}\text{”}\epsilon)$$

  plotted by school

- No noticeable patterns

- Nice set of residuals

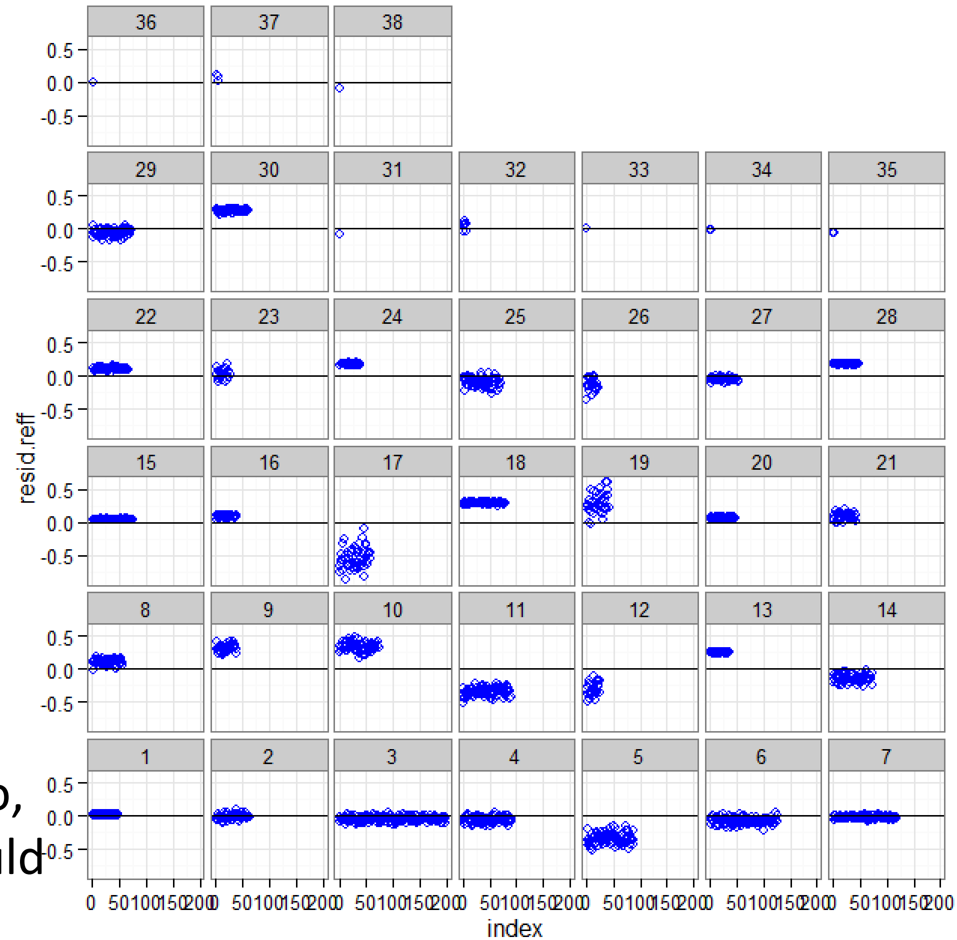


Residuals in the London Schools Data

- Rand Effect residuals

  $$y - X\beta - \epsilon \quad (\text{“=”} Z\eta)$$

  plotted by school
- The *scale* of these residuals is smaller!
- A few schools show noticable deviation from zero
  - We do not expect mean-zero, but the BLUP estimates should cluster around a mean



Uncorrelated Residuals

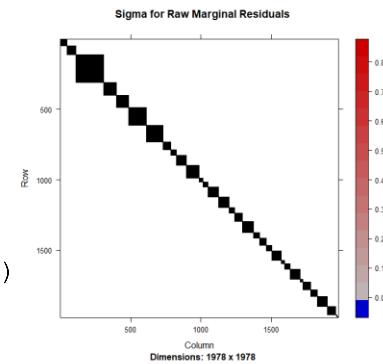
- Correlation in the marginal residuals $y - X\hat{\beta}$ can suggest nonlinearity that isn't really there.
- "***Cholesky residuals***" are marginal residuals, transformed to remove the correlation:

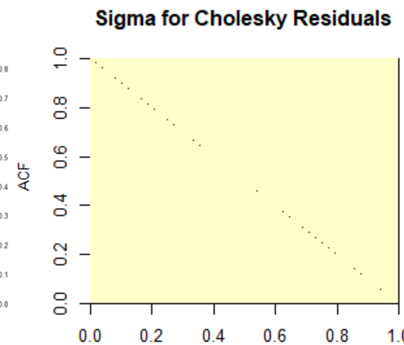
$$
\begin{aligned}
\Sigma = \mathsf{Var}\,(y - X\beta) &= \mathsf{Var}\,(Z\eta + \epsilon) = Z\mathsf{Var}\,(\eta) Z^T + \sigma^2 I \\
SS^T &= \Sigma \quad (\text{i.e. } S = Chol(\Sigma)) \\
e_{chol} &= S^{-1}(y - X\beta)
\end{aligned}
$$

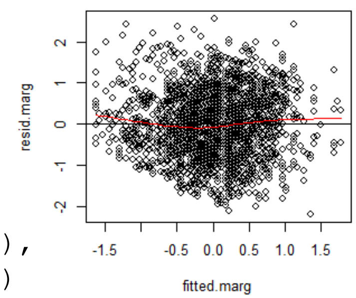
- We use the function getME() to get components of the fitted lmer model to construct Σ, S, and then construct the Cholesky residuals…

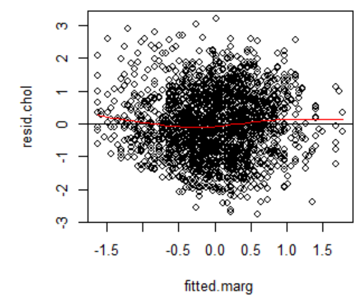
Cholesky Residuals for the London Schools data…

```
> rel.var.eta <- crossprod(getME(lmer.1,"Lambdat"))
> Zt <- getME(lmer.1,"Zt")
> var.epsilon <- sigma(lmer.1)^2
> var.eta <- var.epsilon*(t(Zt) %*% rel.var.eta %*% Zt)
> sI <- var.epsilon *
Diagonal(length(getME(lmer.1,"y")))
> var.y <- var.eta + sI
> S <- chol(var.y)
> resid.chol <- (solve(t(S))%*%resid.marg)@x
>
> fitted.marg <- X%*%beta
>
> image(var.y,main="Sigma for Raw Marginal Residuals")
> image(round(as.matrix(solve(t(S))%*%var.y%*%solve(S)),
+                4), main="Sigma for Cholesky Residuals")
>
> par(mfrow=c(2,2))
> plot(fitted.marg,resid.marg)
> abline(h=0)
> lines(loess.smooth(fitted.marg,resid.marg),col="red")
> plot(fitted.marg,resid.chol)
> abline(h=0)
> lines(loess.smooth(fitted.marg,resid.chol),col="red")
```







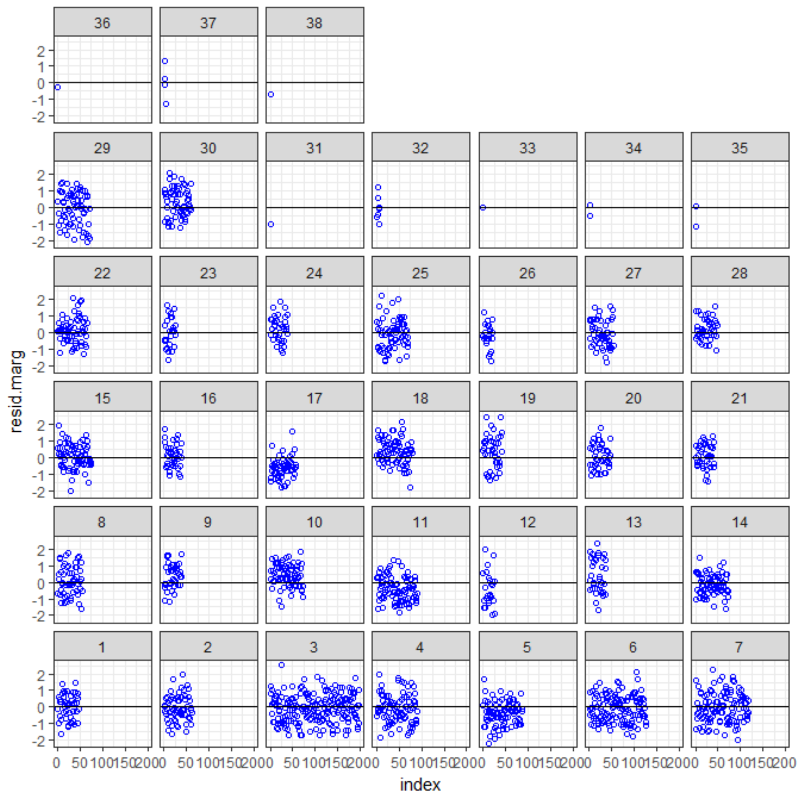


- Not much difference, since correlations are small, in this case…

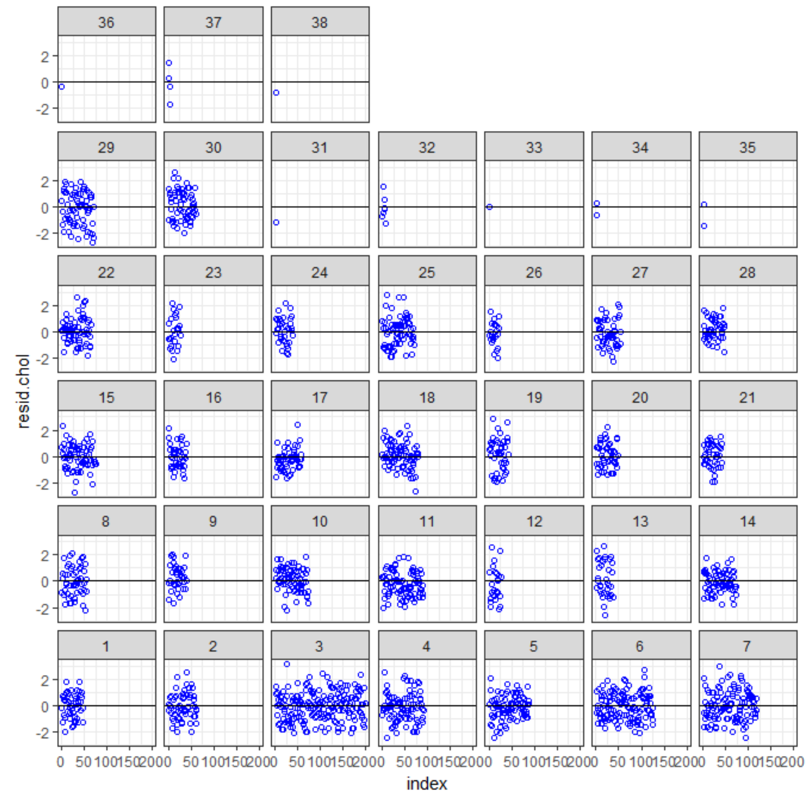
The file “residual-functions.r” provides Cholesky residuals too…

Marginal vs Cholesky Residuals for London Schools Data…

Marginal Residuals



Cholesky Residuals



Very little difference here, since the correlations in the marginal residuals were small to start with…

Standardized Residuals

- We have mostly been looking at raw residuals
- Cholesky Residuals are standardized, so they can be used for outlier detection…
- If you wanted a simpler set of standardized residuals for outlier detection, you could
  - Divide the marginal residuals by the square root of the diagonal elements of Σ = Var(y)
  - Quick and dirty: divide by their sample SD!
  - If there are severe outliers, the sample SD will be inflated and this won't work well

    (unless you omit outliers from the SD calculation)

Residuals: Practical Advice

- Looking at some residuals is better than looking at none.
  - In many MLM’s, marginal and conditional residuals can be used roughly as you would with ordinary linear regression
  - It is worthwhile to plot residuals again the group/cluster indicators
  - To identify and fix problems, plot residuals against other variables (within and/or across clusters), try transformations, etc…
- `residuals(lmer.1)` *gives you the <u>conditional residuals</u>!*

Example: Scheafer's Orthodont Data (Potthoff & Roy, 1964)…

Investigators followed the growth of 27 children (16 males and 11 females). At ages 8, 10, 12 and 14, they measured the distance (in mm) between two points that are easily identified on x-ray exposures of the side of the head. A possible growth curve model is

$$
\begin{aligned}
\text{Dist}_i &= \alpha_{0j[i]} + \alpha_{1j[i]}\text{Age}_i + \epsilon_i, \quad \epsilon_i \overset{indep}{\sim} N(0, \sigma^2) \\
\alpha_{0j} &= \beta_0 + \eta_{0j}, \quad \eta_{0j} \overset{indep}{\sim} N(0, \tau_0^2) \\
\alpha_{1j} &= \beta_1 + \eta_{1j}, \quad \eta_{1j} \overset{indep}{\sim} N(0, \tau_1^2) \\
\text{Cor}(\eta_0, \eta_1) &= \rho
\end{aligned}
$$

i = observation number
j = child number; j[i] = child of i^{th} observation

Example: Sheafer's Orthodontics Data...

```
> data(Orthodont,package="nlme")
> str(Orthodont)
Classes 'nfnGroupedData',
'nfGroupedData', 'groupedData' and
'data.frame':
   108 obs. of  4 variables:
$ distance: num  26 25 29 31 21.5 ...
$ age     : num  8 10 12 14 8 10 ...
$ Subject : Ord.factor w/ 27 levels
  "M16"<"M05"<"M02"<..: 15 15 15 15 3 ...
$ Sex     : Factor w/ 2 levels
  "Male","Female": 1 1 1 1 1 1 1 1 ...

> ## Fit a growth curve model...
> orth.1 <- lmer(distance ~ age +
+ (age | Subject), data=Orthodont)
>
> ## could have got same model as
> ##
> ## orth.1 <- lmer(distance ~ 1 + age +
> ## + (1 + age | Subject), ...)
```

```
> summary(orth.1)
Linear mixed model fit by REML ['lmerMod']
Formula: distance ~ age + (age | Subject)
Data: Orthodont

REML criterion at convergence: 442.6

Scaled residuals:
Min       1Q   Median       3Q      Max
-3.2231 -0.4938  0.0073  0.4722  3.9160

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 Subject  (Intercept) 5.41509  2.3270
age                   0.05127  0.2264 -0.61
Residual              1.71620  1.3100
Number of obs: 108, groups:  Subject, 27

Fixed effects:
            Estimate Std. Error t value
(Intercept) 16.76111    0.77525  21.620
age          0.66019    0.07125   9.265

Correlation of Fixed Effects:
    (Intr)
age -0.848
```

ρ

τ_0^2

τ_1^2

σ^2

β_0

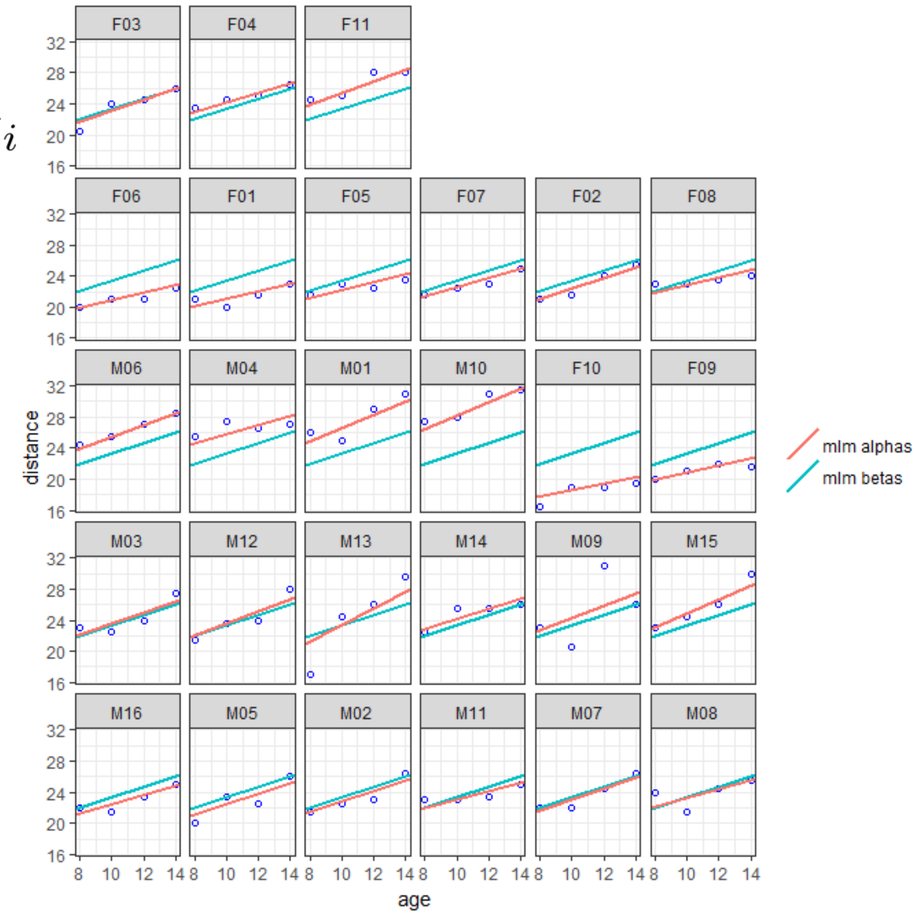
β_1

Orthodont Data: Fitted Growth Curves…

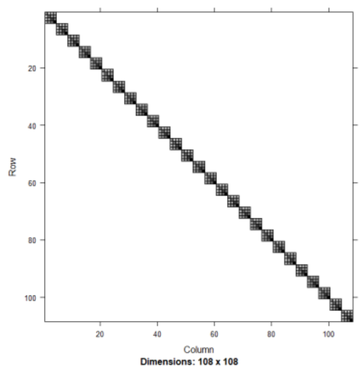
- The model is

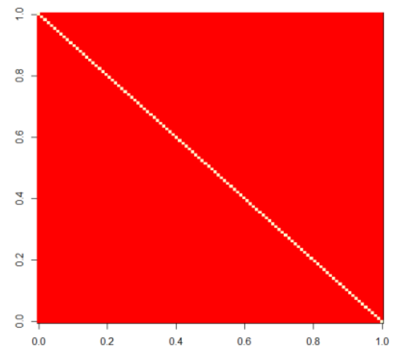
$$\begin{aligned} \text{Dist}_i &= \alpha_{0j[i]} + \alpha_{1j[i]}\text{Age}_i + \epsilon_i \\ \alpha_{0j} &= \beta_0 + \eta_{0j} \\ \alpha_{1j} &= \beta_1 + \eta_{1j} \end{aligned}$$

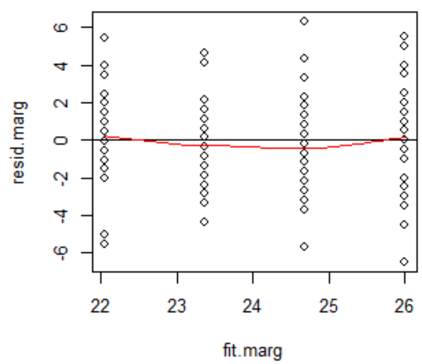
- Green curve (mlm betas) is the overall regression line
- Red curves (mlm alphas) are each individual child’s regression line

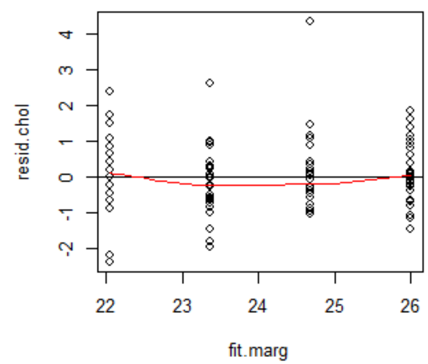


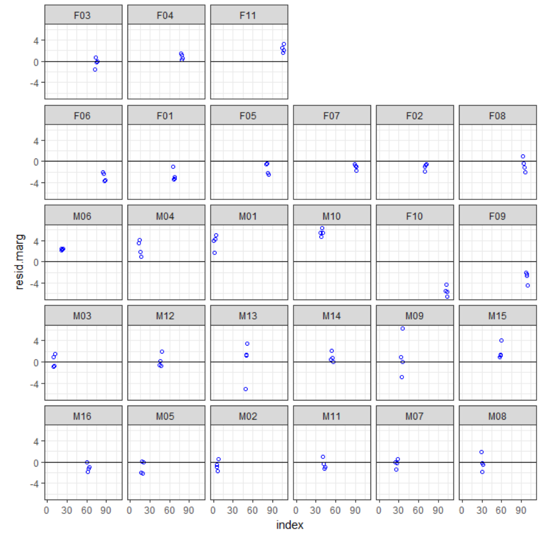
Orthodont: Marginal Residuals

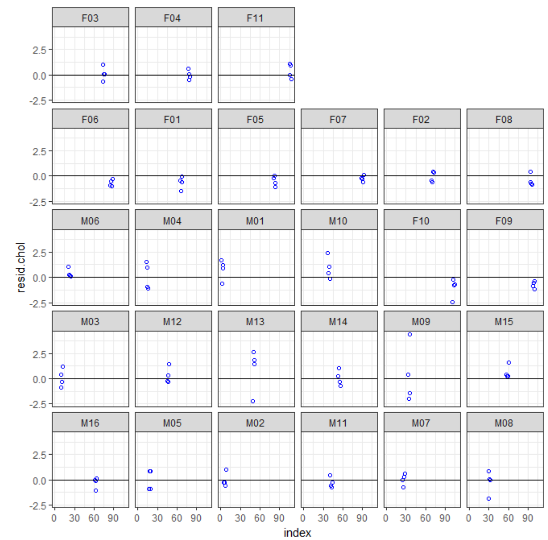




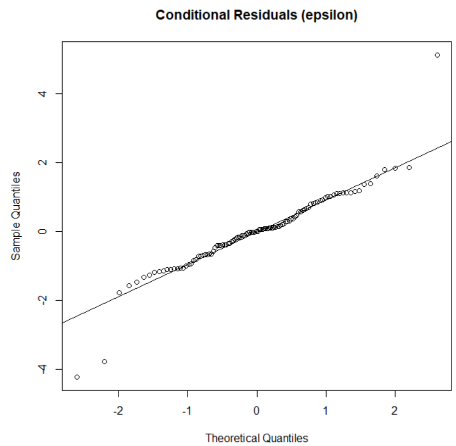


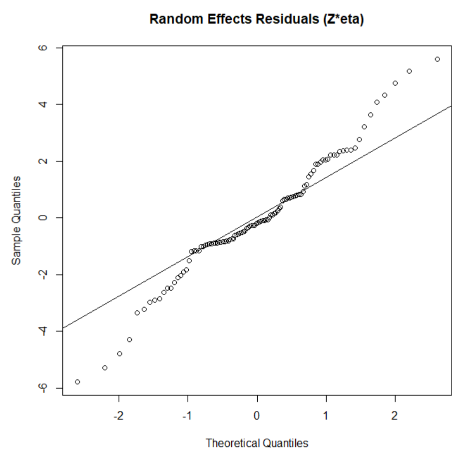


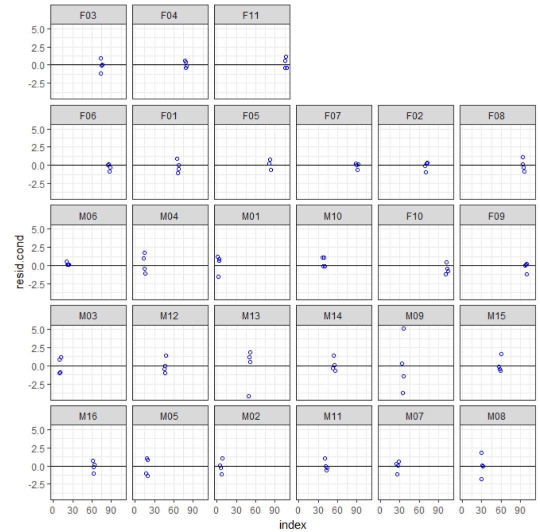


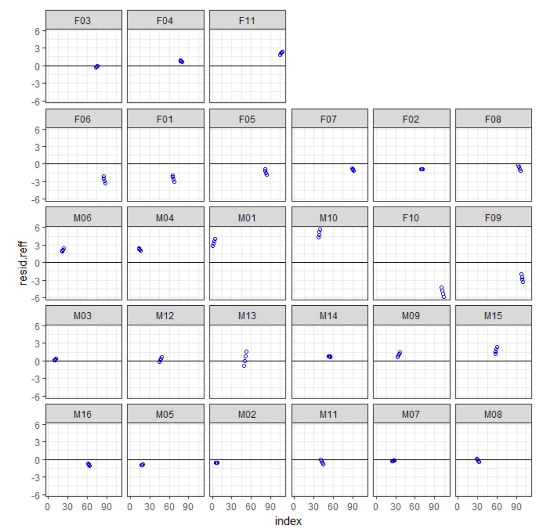


Orthodont: Conditional and Random Effects Residuals









Conclusions

- In General…
  - Marginal and conditional residuals seem quite useful
  - Random effects residuals depend much on structure of $Z\eta$, may not be so useful
  - Standardized or Cholesky residuals useful for outlier detection (maybe also for finding nonlinear xforms!)
- For the London data
  - Our model so far seems ok
  - But we have not yet considered the other covariates in the data set!
- For the Orthodont data
  - Marginal residuals suggest unmodelled male/female differences
  - Conditional residuals seem well-behaved

Summary

- The London Schools Data (again!)
  - A nice random-intercepts, random-slopes model
- Residuals in MLM's
  - Marginal residuals
  - Conditional residuals
  - Random effects residuals
- Cholesky Residuals
- The Orthodont Data