36-617: Applied Linear Models

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

Announcements

- No new reading
- No quiz this week
- HW08 due Weds at 1159pm
- Project description and timeline is¹ available in files area on Canvas

Please read the code and comments in 21 – mlm resduals.r carefully!

Project stuff

- HW09 (see hw09 folder)
 - Many raw materials for Data, Discussion section and (especially)
 Technical Appendix
- Project assignment sheet (see project folder) provides
 - Raw materials for Introduction
 - □ Timeline, guidelines and grading rubric for final IDMRAD paper
- Project Due Dates
 - □ **HW09:** Fri Nov 18 (grace till Sun Nov 20th)
 - Rough IDMRAD draft: Weds Nov 23 (grace till Fri Nov 25th)
 - Peer review: Fri Dec 2 (2 hrs grace!)
 - □ Final IDMRAD paper: Fri Dec 9 (2 hrs grace!)

Office hour this week

Mon and Weds at noon: BJ as usual

- Friday:
 - BJ will take Lorenzo's 11am office hour in 132E Baker (my usual office)
 - Lorenzo will be travelling but will hold a zoom office hour at 3pm Friday. Zoom link TBA.

Outline

- The London Schools Data
 - A nice random-intercepts, random-slopes model
 - ASIDE: Shrinkage for regression lines
- Residuals in MLM's
 - Marginal residuals
 - Conditional residuals
 - Random effects residuals
- Level 1 and Level 2 Residual Plots
- Standardized Residuals
 - ASIDE: Cholesky Residuals
- Practical Advice

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)

We'll focus on the random slopes, random intercepts model...

The MLM is

 $y_{i} = \alpha_{0j[i]} + \alpha_{1j[i]} LRT_{i} + \epsilon_{i}, \quad \epsilon_{i} \sim N(0, \sigma^{2})$ $\alpha_{0j} = \beta_{0} + \eta_{0j}, \quad \eta_{0j} \sim N(0, \tau_{0}^{2})$ $\alpha_{1j} = \beta_{1} + \eta_{1j}, \quad \eta_{1j} \sim N(0, \tau_{1}^{2}), \quad \operatorname{Corr}(\eta_{0j}, \eta_{1j}) = \rho$

with variance components form

 $y_{i} = (\beta_{0} + \beta_{1}LRT_{i}) + (\eta_{0j[i]} + \eta_{1j[i]}LRT_{i}) + \epsilon_{i}$

As an R model this would be Y ~ 1 + LRT + (1 + LRT|school)

ASIDE: Shrinkage for regression lines

```
> ## partially pooled:
> lmer.1 <- lmer(Y ~ 1 + LRT +
+ (1 + LRT | school))
> mlm.alphas <- coef(lmer.1)$school</pre>
>
> ## completely pooled:
> lm.0 < - lm(Y ~ LRT)
> unpooled.betas <- coef(lm.0)</pre>
>
> ## completely unpooled:
> lm.1 < - lm(Y ~ school*LRT - LRT - 1)
> pooled.betas <- coef(lm.1)</pre>
>
> ## "21 - mlm-residuals.r" has
> ## important fitting and
> ## plotting details...
```

The regression lines for the MLM lie between the completely pooled and completely unpooled regression lines; this is the shrinkage phenomenon again



Residuals

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





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}] = α_{oj} + α_{1j} LRT_{ij}??
r_{ii} = y_{ii} - E[y_{ii}]??

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

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



Laird & Ware (1982, *Biometrics*)

- Given the Laird-Ware form $y = X\beta + Z\eta + \epsilon$, can formulate 3 different kinds of residuals:
 - Marginal residuals: $y X\beta$ ("=" $Z\eta + \epsilon$)
 - Conditional residuals:
 - Random effects:

 $y - X\beta - Z\eta \quad (``="\epsilon)$ $y - X\beta - \epsilon \quad (``="Z\eta)$

- In practice, estimate β with $\hat{\beta}$, the MLE, and estimate η with $\eta_{BLUP} \approx E[\eta|$ the data]
 - The "random effects" residuals aren't very useful, but it is good to check the η's themselves!
- Nobre & Singer (2007, Biometrical Journal)

L1: Marginal residuals: $y - X\hat{\beta}$ ("=" $Z\eta + \epsilon$)

- Should be mean 0, but may show grouping structure
- May not be homoskedastic! Will be correlated, unequal variances¹!
- Good for checking fixed effects, just like linear regr.

• <u>L1: Conditional residuals</u>: $y - X\hat{\beta} - Z\hat{\eta}$ ("=" ϵ)

- Should be mean zero with no grouping structure
- Should be homoskedastic!
- Good for checking normality of ϵ , outliers
- <u>L2: Residuals:</u> η
 - Should be mean-zero with no grouping structure
 - Different η 's will have different variances (τ_0^2, τ_1^2 , etc)
 - Good for checking Level 2 normality assumption

```
> row < -1
> str(fixef(lmer.1))
                                               > for (j in 1:J) {
> beta0 <- fixef(lmer.1)[1]</pre>
                                                   col <- 2*j
                                               +
> beta1 <- fixef(lmer.1)[2]</pre>
                                                   nj <- dim(blocks[[j]])[1]</pre>
                                               +
                                                   Z[row: (row+nj-1), c(col-1, col)] <-
                                               +
> str(ranef(lmer.1))
                                                       blocks[[j]]
                                               +
> eta <- ranef(lmer.1)$school</pre>
                                                   row <- row + nj
                                               +
                                               + }
> attach(school.frame)
                                               > beta <- rbind(beta0,beta1)</pre>
> X <- cbind(1, LRT)
                                               > # so beta is a column vector
> blocks <- lapply(split(X, school),</pre>
                                               > eta <- c(t(eta))
    function(x) {matrix(x,ncol=2)})
+
                                               > # so eta is a column vector
> J <- length(blocks)
                                               > resid.marg <- Y - X%*%beta</pre>
> n <- dim(school.frame)[1]</pre>
                                               > resid.cond <- Y - X%*%beta - Z%*%eta
> Z <- matrix(0, nrow=n, ncol=J*2)</pre>
```

11/14/2022

The file "residual-functions.r" provides readable functions to compute these. library(HLMdiag) provides η 's, marginal and conditional residuals automatically...

Marginal residuals

 $y - X \hat{\beta}$ ("=" $Z\eta + \epsilon$) look pretty good...

Conditional residuals
 $y - X\hat{\beta} - Z\hat{\eta}$ ("="\epsilon)
 look pretty good



- Marginal residuals $y - X\hat{\beta}$ ("=" $Z\eta + \epsilon$) plotted by school, vs. $\hat{y}_{marg} = X\hat{\beta}$
- When the η's are larger, the dependence on Z can make these difficult to interpret



Conditional residuals

$$y - X\hat{\beta} - Z\hat{\eta} \quad (``="\epsilon)$$

plotted by school $\hat{y}_{cond} = X\hat{\beta} + Z\hat{\eta}$

- These should not depend on X or Z
- Can be hard to see patterns in facets plot; should also look at ungrouped residuals



Level 1 plots: Normality of conditional residuals ($\hat{\epsilon}$'s)

- > par(mfrow=c(1,1))
- > qqnorm(resid.cond, main=
- + "Conditional Residuals")
- > qqline(resid.cond)
- These Residuals are a little light-tailed relative to the Normal distribution
- But generally they look good



Level 2 plots: Normality of $\hat{\eta}$'s?

```
> eta0 <- ranef(lmer.1)$school[,1]
> eta1 <- ranef(lmer.1)$school[,2]
> par(mfrow=c(2,1))
> qqnorm(eta0,
+ main = "Eta0 (rand. intercepts)")
> qqline(eta0)
> qqnorm(eta1,
+ main = "Eta1 (rand. slopes)")
> qqline(eta1)
```

- The tails seem to be a bit heavier (for both η₀ and η₁) than the normal distribution.
- Generally not bad, for 38 data points (except maybe that low outlier!)

Eta0 (rand. intercepts)



Level 1 plots: $\hat{\epsilon}$'s vs predictors

- > library(HLMdiag)
- > lev.1 <- hlm_resid(lmer.1,</pre>
- + include.ls=F)
- > lev.2 <- hlm_resid(lmer.1,</pre>
- + include.ls=F,level="school")
- > resid.marg <- lev.1\$.mar.resid</pre>
- > resid.cond <- lev.1\$.resid</pre>
- > eta0 <- lev.2\$.ranef.intercept
- > eta1 <- lev.2\$.ranef.lrt</pre>
- > new.data <- data.frame(school,</pre>
- + LRT, resid.marg)
- > ggplot(new.data,
- + aes(x=LRT,y=resid.marg)) +
- + geom_point(aes(color=school)) +
- + geom_smooth()
- > ## and similarly for conditional
 > ## residuals...
- Some evidence of curvature here...



Level 2 plots: $\hat{\eta}$'s vs predictors

- > LRT.avg <- with(school.frame,
- + sapply(split(LRT, school), mean))
- > new.data <- data.frame(LRT.avg,</pre>
- + school=factor(1:38),eta0,eta1)
- > ggplot(new.data,
- + aes(x=LRT.avg,y=eta0)) +
- + geom_point(aes(color=school)) +
- + geom_smooth()
- > ggplot(new.data,
- + aes(x=LRT.avg,y=eta1)) +
- + geom_point(aes(color=school)) +
- + geom_smooth()
- Evidence of curvature, especially in η_1
- Suggests a transformation of LRT, perhaps a cubic?



Standardized Residuals

- There can be two "standardization" problems with Level 1 MLM residuals
 - Like lm() residuals, they have <u>unequal variance</u>
 - Raw residuals for Im() have variances like $\hat{\sigma}^2 \sqrt{1 h_{ii}}$
 - Similar but more complicated for $\hat{\epsilon}_{marg}$ and $\hat{\epsilon}_{cond}$
 - $\hat{\epsilon}_{marg}$ will have noticeable correlations induced by Z η
 - $\hat{\epsilon}_{cond}$ should be approximately uncorrelated
- Level 2 residuals, $\hat{\eta}$'s, should be approximately uncorrelated but also have unequal variances

Calculating standardized residuals

- These can be calculated¹ "by hand" but it is easier to use library(HLMdiag):
- > h.stdres.1 <- hlm_resid(lmer.1,level=1,include.ls=F,standardize=T)</pre>



¹See demonstration at end of "21 – mlm residuals.r" ₂₃

ASIDE: Cholesky Residuals

- Correlation in the marginal residuals $y X\hat{\beta}$ can make it hard to interpret qqnorm plots, etc.
- "Cholesky residuals" are marginal residuals, transformed to remove the correlation:

$$\Sigma = \operatorname{Var}(y - X\beta) = \operatorname{Var}(Z\eta + \epsilon) = Z\operatorname{Var}(\eta)Z^T + \sigma^2 I$$

$$SS^T = \Sigma \text{ (i.e. } S = Chol(\Sigma)\text{)}$$

$$e_{chol} = S^{-1}(y - X\beta)$$

- In the R file accompanying this lecture we show how to get components of the fitted Imer model to construct Σ, S, to construct Cholesky residuals...
- Using hlm_resid() from library(HLMdiag) easier!

Level 1: Standardized vs Not

- > sresid.cond <-
- + h.stdres.1\$.std.resid
- > cresid.marg <-</pre>
- + h.stdres.1\$.chol.mar.resid
- > qqnorm(resid.cond, main=
- + "Conditional Residuals")
- > qqline(resid.cond)
- > qqnorm(sresid.cond, main=
- + "Stdized Cond Residuals")
- > qqline(sresid.cond)
- > qqnorm(resid.marg, main=
- + "Marginal Residuals")
- > qqline(resid.marg)
- > qqnorm(cresid.marg, main=
- + "Cholesky Marg Residuals")
- > qqline(cresid.marg)



• Level 1 Standardization makes little difference for this data

Level 2: Standardized vs Not





 Level 2 Standardization slightly improves normality

Residuals: Practical Advice

- Looking at some residuals is better than looking at none.
 - In many MLM's, <u>marginal</u> and <u>conditional residuals</u> can be used roughly as you would with ordinary linear regression
 - Good to look at facet plots of Level 1 residuals
 - Some problems will be easier to see with ungrouped plots of Level 1 (ê marg and ê cond) and Level 2 (ŋ's) residual plots (vs. ŷ's, predictor variables, etc.)
 - Standardized residuals helpful for assessing normality, outliers
- If you forget HLMdiag, residuals (lmer.1) gives you the conditional residuals

Summary

- The London Schools Data
 - A nice random-intercepts, random-slopes model
 - ASIDE: Shrinkage for regression lines
- Residuals in MLM's
 - Marginal residuals
 - Conditional residuals
 - Random effects residuals
- Level 1 and Level 2 Residual Plots
- Standardized Residuals
 - ASIDE: Cholesky Residuals
- Practical Advice