36-617 Applied Linear Models

Lmer estimation and model selection

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 is[1] available in files area on Canvas

- Please read the code and comments in **`22 - estimation and model selection.r`** carefully!

[1](…or will be, by Friday of Week 10!)

Project stuff

- HW09 (see hw09 folder)
  - Many raw materials for **Data, Discussion** sections 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 ***4pm*** Friday. Zoom link: https://cmu.zoom.us/j/98282667112?pwd=dm8yWXR5NkdFemRlbnFYeXBEQXVVQT09

Outline

- Marginal and Conditional Models
- Estimation
  - MLE: Full maximum likelihood
  - EB: Empirical Bayes
  - REML: Restricted or Residual maximum likelihood
- Likelihood Ratio Tests, AIC, BIC
  - Change from REML to MLE for all three
- Df for marginal and conditional models
  - DIC, cAIC
- Variable selection: Practical Advice
- Example (London Schools)

Marginal and Conditional Models

Consider the general Laird-Ware formulation (I'm being a little careful to underline things that are vectors...)

$$\begin{aligned} \underline{Y} &= X\underline{\beta} + Z\underline{\eta} + \underline{\varepsilon} \\ \underline{\varepsilon} &\sim N(\underline{0}, \sigma^2 I) \\ \underline{\eta} &\sim N(\underline{0}, \Psi) \end{aligned}$$

The variance-covariance matrix $\sigma^2 I$ and the variance-covariance matrix Ψ are both pretty big, but also pretty simple. For example in the random slopes / random intercepts model,

$$\sigma^2 I = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \cdots & 0 \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} \quad \Psi = \begin{bmatrix} \tau_0^2 & \rho\tau_0\tau_1 & 0 & 0 & \cdots & 0 \\ \rho\tau_0\tau_1 & \tau_1^2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \tau_0^2 & \rho\tau_0\tau_1 & \cdots & 0 \\ 0 & 0 & \rho\tau_0\tau_1 & \tau_1^2 & \cdots & 0 \\ \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 0 & \tau_0^2 & \rho\tau_0\tau_1 \\ 0 & 0 & \cdots & 0 & \rho\tau_0\tau_1 & \tau_1^2 \end{bmatrix}$$

and in general, $\Psi = \Psi(\underline{\omega})$, where $\underline{\omega}$ = (a few τ^2's and a few ρ's).

Marginal and Conditional Models

The Laird-Ware formulation

$$\begin{aligned} \underline{Y} &= X\underline{\beta} + Z\underline{\eta} + \underline{\varepsilon} \\ \underline{\varepsilon} &\sim N(\underline{0}, \sigma^2 I) \\ \underline{\eta} &\sim N(\underline{0}, \Psi) \end{aligned}$$

tells us that the *joint pdf* for $\underline{Y}$ and $\underline{\eta}$, given σ^2 and $\underline{\omega} =$ (a few τ^2's and a few ρ's) is

$$f(\underline{Y}, \underline{\eta} | \underline{\beta}, \sigma^2, \underline{\omega}) = \textit{(Some multivariate normal density)}$$

We can figure out a lot about this density by factoring it

$$f(\underline{Y}, \underline{\eta} | \underline{\beta}, \sigma^2, \underline{\omega}) = f(\underline{Y} | \underline{\beta}, \underline{\eta}, \sigma^2) f(\underline{\eta} | \underline{\omega})$$

(abusing notation to use the "f" to refer to different densities).

- $f(\underline{Y} | \underline{\beta}, \underline{\eta}, \sigma^2)$ is the *conditional model*
- $f(\underline{Y} | \underline{\beta}, \sigma^2, \underline{\omega}) = \int f(\underline{Y} | \underline{\beta}, \underline{\eta}, \sigma^2) f(\underline{\eta} | \underline{\beta}, \underline{\omega})\, d\underline{\eta}$ is the *marginal model*

These correspond to the conditional and marginal residuals we looked at last time!

Marginal and Conditional Models

With the Laird-Ware formulation

$$\begin{aligned} \underline{Y} &= X\underline{\beta} + Z\underline{\eta} + \underline{\varepsilon} \\ \underline{\varepsilon} &\sim N(\underline{0}, \sigma^2 I) \\ \underline{\eta} &\sim N(\underline{0}, \Psi) \end{aligned}$$

we can see that

- The conditional model $f(\underline{Y}|\underline{\beta}, \underline{\eta}, \sigma^2)$ specifies that

$$(\underline{Y}|\underline{\beta}, \underline{\eta}, \sigma^2) \sim N(X\underline{\beta} + Z\underline{\eta}, \sigma^2 I)$$

- The marginal model $f(\underline{Y}|\underline{\beta}, \sigma^2, \underline{\omega})$ specifies that

$$(\underline{Y}|\underline{\beta}, \sigma^2, \underline{\omega}) \sim N(X\underline{\beta}, \mathsf{Var}(Z\eta + \varepsilon) = N(X\underline{\beta}, Z\Psi(\omega)Z^T + \sigma^2 I)$$

Estimation: Maximum Likelihood

Maximum Likelihood estimates are obtained from the marginal model

$$\begin{aligned} Y &\sim N(X\beta, \Sigma(\omega, \sigma^2)) \\ \text{where} \quad \Sigma(\underline{\omega}, \sigma^2) &= Z\Psi(\omega)Z^T + \sigma^2 I \end{aligned}$$

which means that the likelihood for Y is[1]

$$(2\pi)^{-n/2}|\Sigma(\underline{\omega}, \sigma^2)|^{-1/2} \exp\left(-\frac{1}{2}(\underline{y} - X\underline{\beta})^T \Sigma(\underline{\omega}, \sigma^2)^{-1}(\underline{y} - X\underline{\beta})\right)$$

so $-2\log($*likelihood*$)$ is (up to an additive constant that we can ignore)

$$(\underline{Y} - X\underline{\beta})^T \Sigma^{-1}(\underline{\omega}, \sigma^2)(\underline{Y} - X\underline{\beta}) + \log|\Sigma(\underline{\omega}, \sigma^2)| \qquad (*)$$

If we can minimize $(*)$ with respect to $\underline{\beta}, \underline{\omega}, \sigma^2$, we will have maximized the likelihood, and the values of the parameters that we find will be the MLE's $\hat{\underline{\beta}}, \hat{\underline{\omega}}, \hat{\sigma}^2$.

[1]Here we define |A| = det(A).

Estimation: Maximum Likelihood

To minimize

$$(\underline{Y} - X\underline{\beta})^T \Sigma^{-1}(\underline{\omega}, \sigma^2)(\underline{Y} - X\underline{\beta}) + \log |\Sigma(\underline{\omega}, \sigma^2)| \qquad (*)$$

with respect to $\underline{\beta}, \underline{\omega}, \sigma^2$ we want to (schematically, anyway):

- Get an initial estimate for $\hat{\underline{\beta}}$, perhaps just the ordinary least-squares $\hat{\beta}$'s.
- Iterate the following three steps[2] until changes in the values of $(*)$ are less than some "tolerance" (like 0.0002 or something):
  1. Plug our current estimate for $\hat{\underline{\beta}}$ into $(*)$, and then minimize $(*)$ with respect to $\underline{\omega}$ and σ^2 to get new estimates $\hat{\underline{\omega}}, \hat{\sigma}^2$.
  2. Use these new values for $\underline{\omega}$ and $\hat{\sigma}^2$ to re-calculate $\Sigma(\hat{\underline{\omega}}, \hat{\sigma}^2)$, plug the new value into $(*)$ and use the method of Generalized Least Squares (GLS) to obtain new estimates for $\hat{\beta}$.
  3. Evaluate $(*)$ at the new estimates of $\hat{\underline{\beta}}, \hat{\underline{\omega}}, \hat{\sigma}^2$.

Steps 2 and 3 are generally well-solved problems in numerical analysis. On the other hand, Step 1 can be quite difficult. That is part of the reason R makes 6 or 7 optimizers available for estimating MLM's.

[2](an example of a Gauss-Seidel algorithm)

Estimation: Empirical Bayes

- After the MLE's $\underline{\hat{\beta}}$, $\hat{\sigma}^2$, and $\underline{\widehat{\omega}} =$ (some $\hat{\tau}^2$'s & some $\hat{\rho}$'s) have been obtained, we can plug them into the calculation

$$\underline{\hat{\eta}}_{EB} = E\left[\underline{\eta} \middle| \underline{Y}, \underline{\hat{\beta}}, \hat{\sigma}^2, \underline{\widehat{\omega}}\right] = \int \underline{\eta} f\left(\underline{\eta} | \underline{Y}, \underline{\hat{\beta}}, \hat{\sigma}^2, \underline{\widehat{\omega}}\right) \mathrm{d}\underline{\eta}$$

- The density $f\left(\underline{\eta} | \underline{Y}, \underline{\hat{\beta}}, \hat{\sigma}^2, \underline{\widehat{\omega}}\right)$ can be calculated from the other conditional and marginal densities, using Bayes' rule.
- The point estimates $\underline{\hat{\eta}}_{EB}$ are called *empirical Bayes* estimates.
  - Standard deviations for $\underline{\hat{\eta}}_{EB}$ can be obtained in the same way.
  - There is a lot of numerical analysis involved in evaluating the integral.
- The same thing is done with REML parameter estimates…

Estimation: Maximum Likelihood

- Pro's:
  - $\underline{\hat{\beta}}_{MLE}$ is unbiased, just like in ordinary regression.
  - MLE's are needed for model comparisons
- Con's:
  - $\hat{\sigma}^2_{MLE}$ is biased, just like ordinary regression. Estimates of the τ^2's and ρ's in $\underline{\widehat{\omega}}$ are biased as well.
  - For larger models, the full Maximum Likelihood procedure can be slow.
  - Difficult to use standard tools to test $\tau^2 = 0$, e.g.

Estimation: REML

- *Restricted* (or residual, or reduced) *maximum likelihood (REML)* is based on the same idea as estimating σ^2 in ordinary regression:
  - If $\underline{Y} = X\underline{\beta} + \underline{\epsilon}$, then for $H = X\left(X^T X\right)^{-1} X^T$,

$$(I-H)\underline{Y} = (I-H)\left(X\underline{\beta} + \underline{\epsilon}\right)$$
$$= (I-H)\underline{\epsilon} \sim N\left(0, (I-H)\sigma^2\right)$$

  - Then we get a $\underline{\beta}$-free, unbiased estimate of σ^2 as

$$\hat{\sigma}^2 = \frac{1}{n-k} RSS = \frac{1}{n-k}\left[(\mathrm{I}-\mathrm{H})\underline{Y}\right]^T \left[(\mathrm{I}-\mathrm{H})\underline{Y}\right]$$

- REML applies the same idea to MLM's

Estimation: REML

For the MLM $\underline{Y} = X\underline{\beta} + Z\underline{\eta} + \underline{\epsilon}$, if we take $H = X(X^TX)^{-1}X^T$ again, we get that

$$\begin{aligned}(I-H)\underline{Y} &= (I-H)(X\underline{\beta} + Z\underline{\eta} + \underline{\epsilon}) = (I-H)Z\underline{\eta} + (I-H)\underline{\epsilon} \\ &\sim N(0,\ (I-H)Z\Sigma(\underline{\omega}, \sigma^2)Z^T(I-H))\end{aligned}$$

- We can use this fact to obtain obtain β-free estimates $\hat{\underline{\omega}}_{REML}$ and $\hat{\sigma}^2_{REML}$. We still have a challenging optimization problem, but it only has to be done once (no iteration back and forth with $\hat{\underline{\beta}}$).

- Finally we obtain $\hat{\underline{\beta}}_{REML}$ estimates using a GLS procedure similar to what was done with $(*)$.

It can be shown that:

- $\Sigma(\hat{\underline{\omega}}_{MLE}, \hat{\sigma}^2_{MLE})$ is biased, but $\Sigma(\hat{\underline{\omega}}_{REML}, \hat{\sigma}^2_{REML})$ is unbiased
- $\hat{\beta}_{REML}$ and $\Sigma(\hat{\underline{\omega}}_{REML}, \hat{\sigma}^2_{REML})$ are not maximum likelihood estimates

Since people are often more interested in the random effects than the fixed effects in MLM's (and because it's fast), REML is often the default estimation method.

Likelihood Ratio Tests

- If M0 is nested in M1 *(obtain M0 from M1 by making linear restrictions on the parameters),* and if the data came from M0, then as the sample size grows

$$-2[\log(likelhood_{M0}) - \log(likelihood_{M1})]$$

will be distributed as χ^2 on k df, where k is the number of linear restrictions (usually, difference in # of parameters).

- Cautions
  - Evaluate $\log(likelihood)$ at the MLE's for the model
  - If M0 has parameter values at the edge of the parameter space for M1, LRT may not be chi-squared[1] under H_0.

[1]Self, S. G., & Liang, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, *82*(398), 605-610.

Likelihood Ratio Tests

- Because the likelihoods should be evaluated at the MLE's, it is better to use MLE for estimation than REML.
  - REML estimates can be "close enough", but for any particular problem we don't know if they'll be close enough unless we calculate the MLE's too (in which case, just use the MLE's!)

- If M0 has parameter value(s) at the edge of the parameter space for M1, the LRT may not have a chi-squared distribution[1] under M0. So:
  - LRT fine for restrictions on β's or ρ's
  - LRT Not OK for testing whether $\tau^2 = 0$, or equivalently, comparing a model with or without a particular random effect.
    (LRT tends to be conservative: choosing $\tau^2 = 0$ when $\tau^2 > 0$)
- In MLM's counting df can be tricky – see slides on AIC, BIC, etc.

[1]Self, S. G., & Liang, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, *82*(398), 605-610.

Information Criteria: AIC & BIC

- AIC = -2logLik(M) + 2 k
  - Tends to pick models with lower prediction error
- BIC = -2logLik(M) + k log(n)
  - Tends to pick models closer to the “correct” model
- Issues:
  - BIC picks simpler models than AIC
    - Models do not have to be nested
  - Need logLik(M) evaluated at, or very near, MLE’s
    - REML estimates often not good enough[1]
  - What should k (degrees of freedom) be?

[1]Can use update(…, REML=F) or lmer(…,REML=F) to force MLEs…

k=df for Marginal Model

- Marginal model:

$$Y|\underline{\beta},\underline{\omega},\sigma^2 \sim N(X\underline{\beta},\Sigma(\underline{\omega},\sigma^2))$$

- Degrees of freedom are straightforward:

  k = (number of β's in $\underline{\beta}$) + (number of τ^2's & ρ's in $\underline{\omega}$)

- R's AIC, BIC and logLik functions:
  - Log(likelihood) based on marginal model $f(\underline{Y}\,|\underline{\beta},\underline{\omega},\sigma^2)$
  - df as above[1] ; remember to set REML=F
  - Fine for testing β's or ρ's (0 is in the middle of param space)
  - Not great for testing η's or τ^2's (0 is at the boundary)

[1]R might use this df + 1 (for σ^2). But it would not matter for AIC or BIC, since the +1 would cancel out when we compare different models.

k=df for Conditional Model

- Conditional model:

$$Y|\underline{\eta},\underline{\beta},\underline{\omega},\sigma^2 \sim N(X\underline{\beta} + Z\underline{\eta},\sigma^2 I)$$

- Degrees of freedom (?):

k = (number of β's in $\underline{\beta}$) + (number of η's in $\underline{\eta}$?)

- The appropriate df for $\underline{\eta}$ not so obvious…

$$y_i = \alpha_{j[i]} + \epsilon_i, \; \epsilon_i \sim N(0,\sigma^2)$$

$$\alpha_j = \beta_0 + \eta_j, \; \eta_j \sim N(0,\tau_0^2)$$

  - τ_0^2 large $\Rightarrow$ one-way ANOVA with J cells (df=J)
  - τ_0^2 small $\Rightarrow$ fitting grand mean only (df=1)
  - $1 \leq k_{eff} \leq J$, depending on size of τ_0^2

How to select random effects??

- Many schools of thought but I will briefly discuss just two: DIC and cAIC.
- Both are modifications of AIC
- DIC = -2logLik(M) + 2 k_{eff}
  - LogLik(M) based on marginal model $f(\underline{Y} \mid \underline{\beta}, \underline{\omega}, \sigma^2)$
  - k_{eff} is estimated from the curvature of the likelihood, which is driven by the size of the τ^2's
- cAIC = -2logLik(M) + 2 k_{eff}
  - logLik(M) based on the conditional model $f(\underline{Y} \mid \underline{\eta}, \underline{\beta}, \underline{\omega}, \sigma^2)$
  - k_{eff} is a different "model curvature" estimate

Variable Selection: Practical Advice

- Start with multilevel model that represents your initial guesses about group structure in the data
- Selection on all the fixed effects first, using AIC or BIC
  - AIC will result in ***bigger models*** that predict better
  - BIC will result in ***smaller models*** that interpret better
- Then use DIC or cAIC to select random effects
  - *True confessions*:
    - Others find cAIC useful; for me, DIC is usually enough.
    - For informal explorations I am occasionally lazy and use AIC or BIC for η's – *not great!*
  - We will talk later in the semester about other, simulation-based methods for selecting random effects.

Variable Selection: Practical Advice

- Whenever you put an interaction in a model, you should also put the lower order terms in the model (R usually does this for you)
  - X*Y expands to 1 + X + Y + X:Y
  - X*Y*Z expands to 1 + X + Y + Z + X:Y + X:Z + Y:Z + X:Y:Z   (etc. etc.)
  - Similarly for polynomials: if you put I(X^3) in a model, make sure 1 + X + I(X^2) + I(X^3) are in the model *(R doesn't do this for you!)*
- Whenever you put a random effect in a model, include the same term as a fixed effect
  - If you want (1 + X + Y|group), make sure the model includes 1 + X + Y + (1+X+Y|group)
  - It's OK to have fixed effects that are not also random effects
- Like all rules, there are times that these should be broken

Example: London Schools Data

```
## adding two variables to the school.frame data
> school.frame$sch.avg <- with(school.frame, unlist(sapply(split(LRT,school),
+                          function(x) {rep(mean(x),length(x))})))
> school.frame$LRT.cwc <- with(school.frame,LRT-sch.avg)
> str(school.frame)
'data.frame':   1978 obs. of  9 variables:
 $ Y            : L1: End-of-year test score
 $ school       : L2: School ID
 $ LRT          : L1: Beginning-of-year score
 $ Gender       : L1: Female or male
 $ School.denom : L2: Factor w/ 4 levels "CofE","Other","RomCath", "State"
 $ School.gender: L2: Factor w/ 3 levels "All.Boy","All.Girl", "Mixed"
 $ VR           : L1: Factor w/ 3 levels "High","Low","Med"
 $ sch.avg      : L2: LRT cluster (school) means
 $ LRT.cwc      : L1: LRT, centered within cluster (school)
> ##
> ## It’s good to make a note of which variables vary with individuals
> ## (students; L1), and which ones vary with cluster (school; L2).
```


Source: Goldstein (1993); see also library(mlmREv)

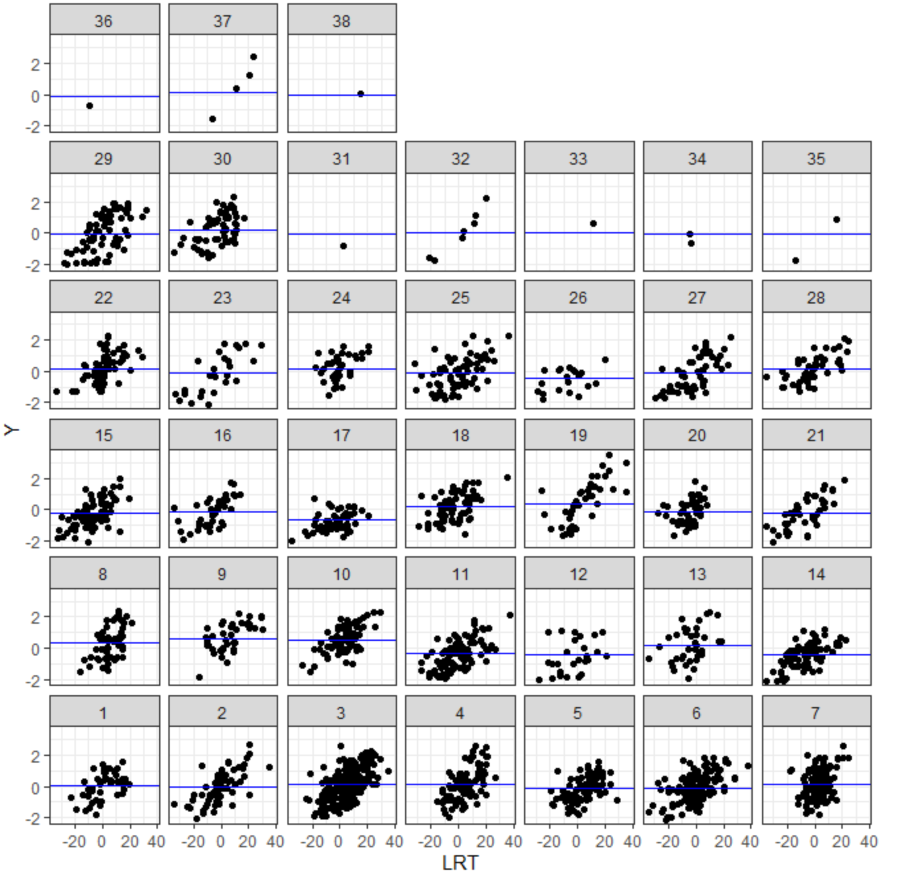
On the next pages, we'll quickly compare six models

- Random Intercept Only
  - Y ~ 1 + (1 | school)
- Random Intercept, Fixed Slope
  - Y ~ LRT + (1 | school)
- Random Intercept, Random Slope
  - Y ~ LRT + (LRT | school)
- Random Intercept, Random Slope, CWC
  - Y ~ LRT.cwc + (1 + LRT.cwc | school)
- Random Intercept, individual and group predictor
  - Y ~ LRT + sch.avg + (1 + LRT | school)
- Random Intercept, Cross-Level Interaction
  - Y ~ LRT * sch.avg + (1 + LRT | school)

Details in school-frame-mlm.r

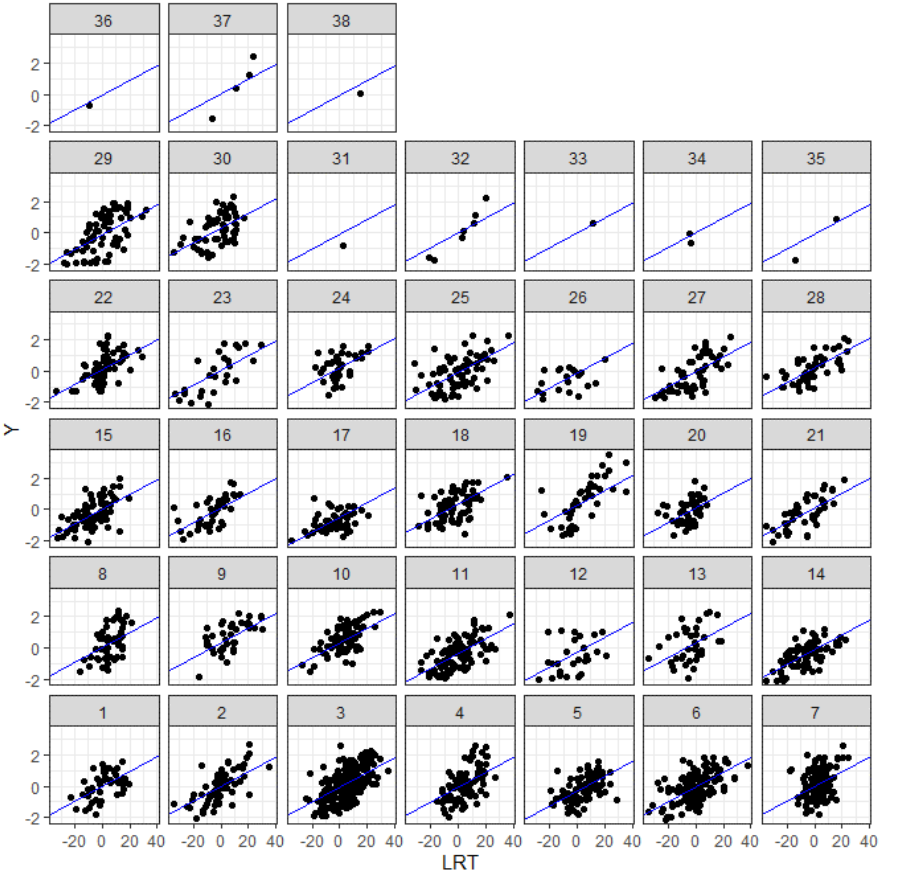
Random Intercept Only

```
> lmer.1 <- lmer(Y ~ 1 +
+ (1 | school), data=school.frame)
> display(lmer.1)
lmer(formula = Y ~ 1 + (1 | school),
data = school.frame)
coef.est  coef.se
   -0.01     0.06

Error terms:
 Groups   Name        Std.Dev.
 school   (Intercept) 0.30
 Residual             0.96
```



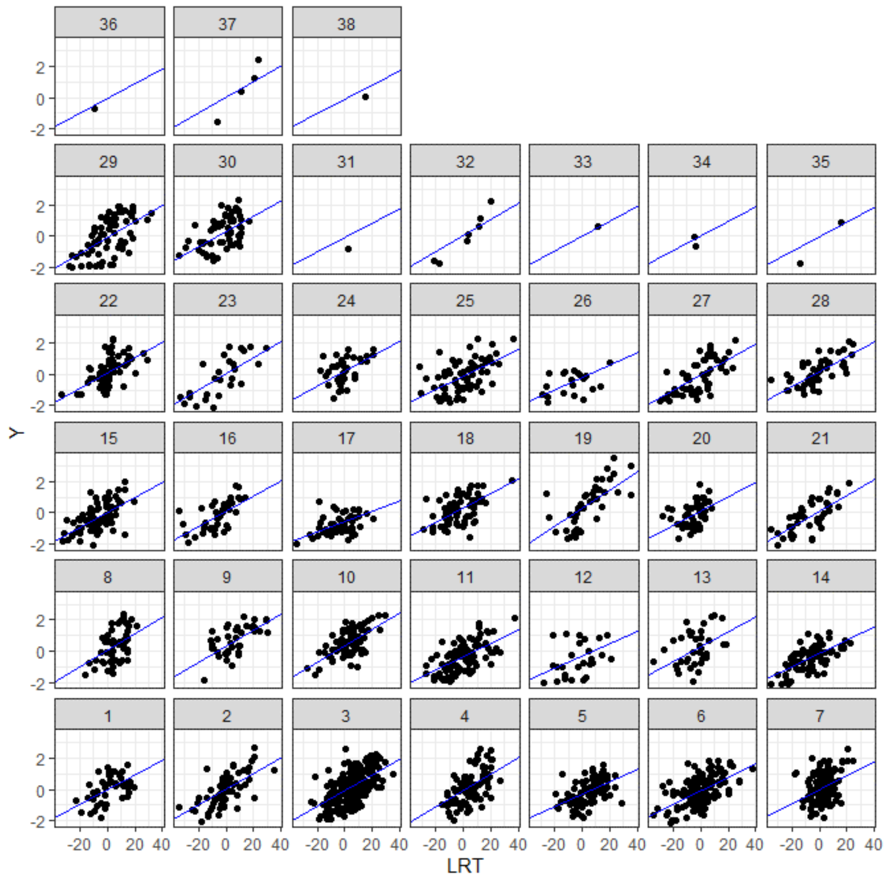
Random Intercept, Fixed Slope

```
> lmer.2 <- lmer(Y ~ LRT +
+ (1 | school), data=school.frame)
> display(lmer.2)
lmer(formula = Y ~ LRT + (1 |
school), data = school.frame)
            coef.est coef.se
(Intercept) 0.01     0.05
LRT         0.05     0.00

Error terms:
 Groups   Name        Std.Dev.
 school   (Intercept) 0.23
 Residual             0.79
```



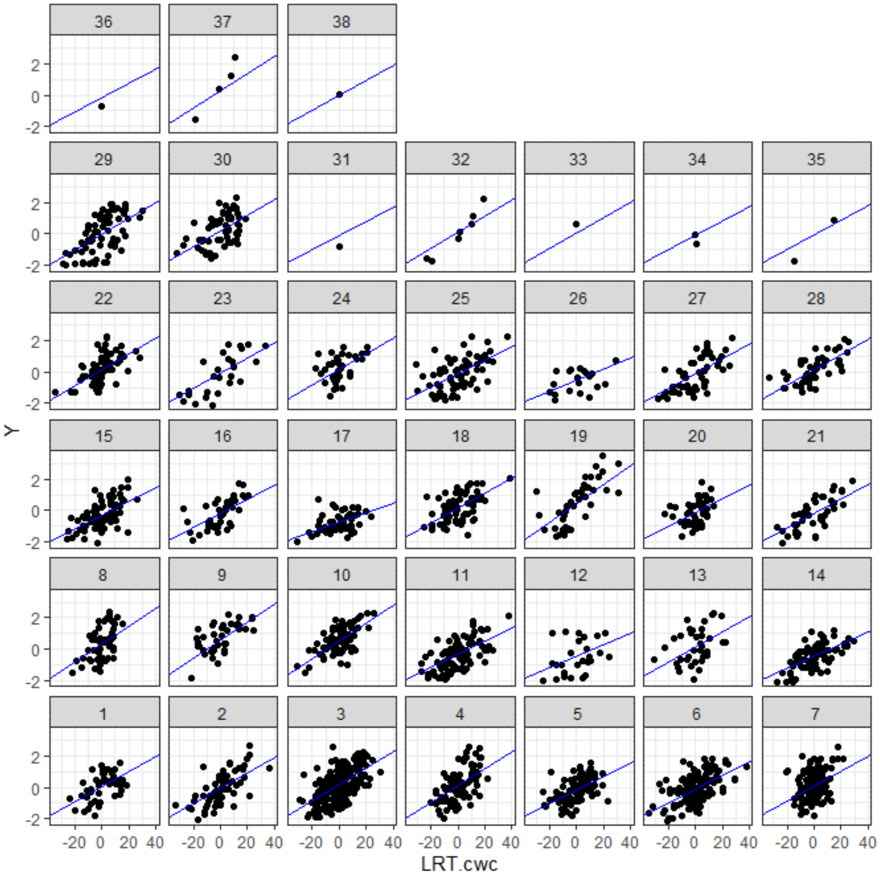
Random Intercept, Random Slope

```
> lmer.3 <- lmer(Y ~ LRT +
+ (LRT | school), data=school.frame)
 Model failed to converge with
max|grad| = 0.235913 (tol = 0.002)
> display(lmer.3)
lmer(formula = Y ~ LRT + (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   (Intercept) 0.24
          LRT         0.01     0.57
 Residual             0.79
```



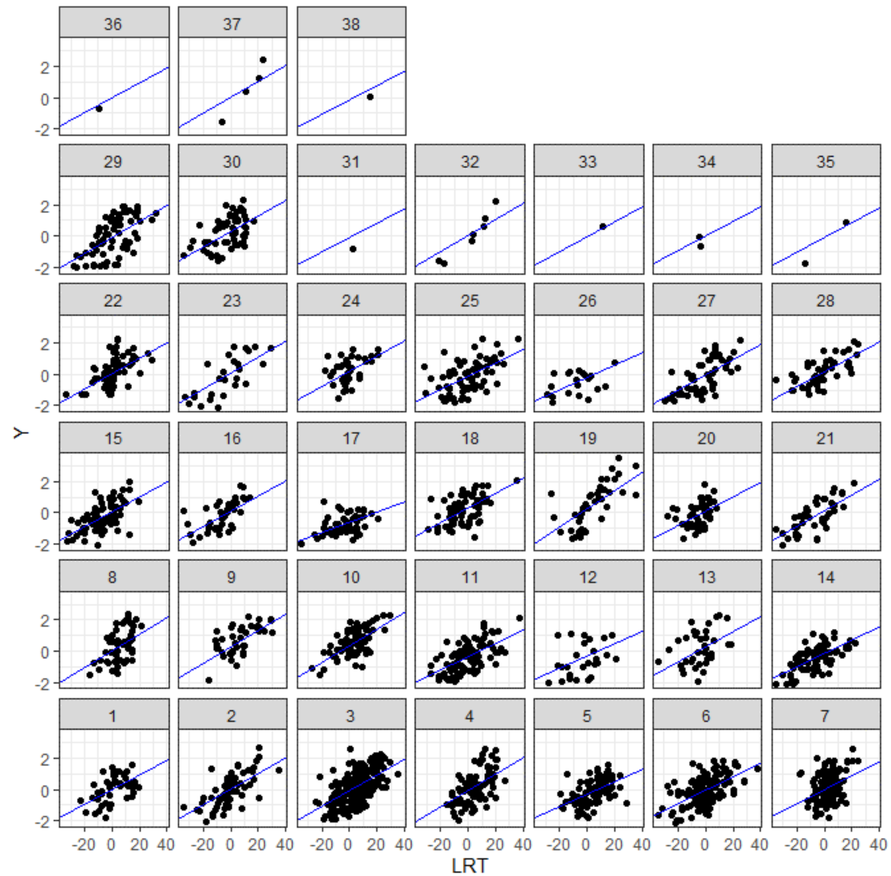
Random Intercept, Random Slope, CWC

```
> lmer.4 <- lmer(Y ~ LRT.cwc +
+ (1 + LRT.cwc | school),
+ data=school.frame)
Model failed to converge with
max|grad| = 0.0042402 (tol = 0.002)
> display(lmer.4)
lmer(formula = Y ~ LRT.cwc + (1 +
LRT.cwc | school), data =
school.frame)
            coef.est coef.se
(Intercept) -0.01     0.06
LRT.cwc      0.05     0.00

Error terms:
 Groups   Name        Std.Dev. Corr
 school   (Intercept) 0.32
          LRT.cwc     0.01     0.83
 Residual             0.79
```



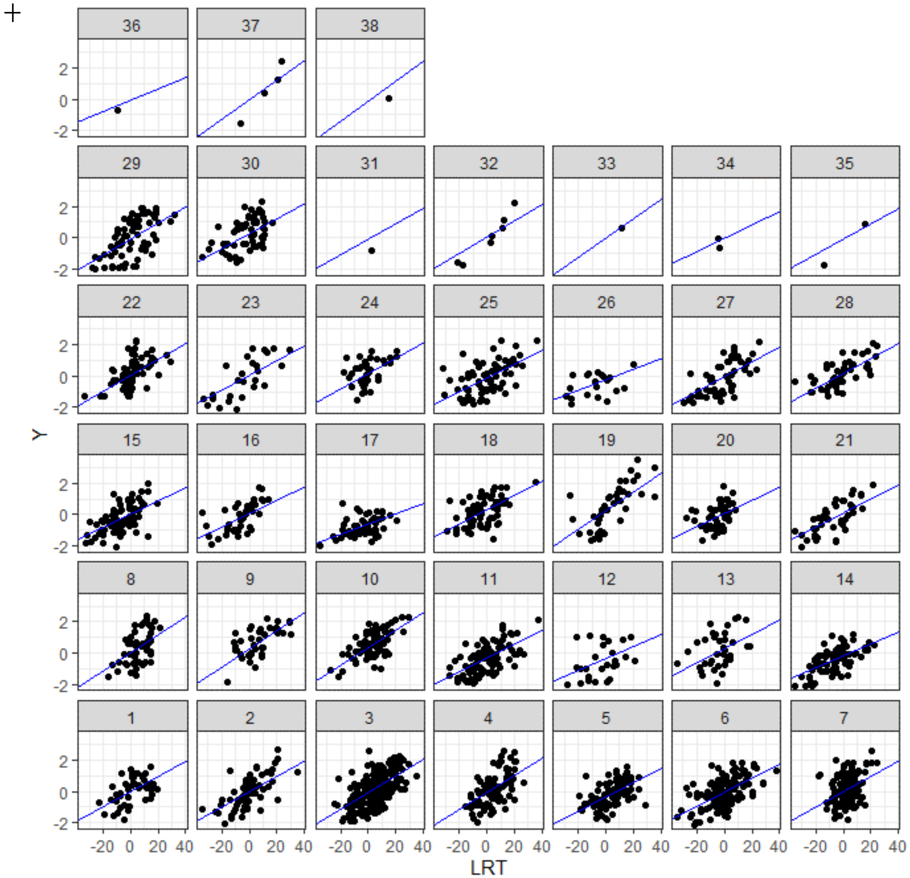
Random Intercept, individual and group predictor

```
> lmer.5 <- lmer(Y ~ LRT +
+ sch.avg + (1 + LRT | school),
+  data=school.frame)
> display(lmer.5)
lmer(formula = Y ~ LRT + sch.avg +
(1 + LRT | school), data =
school.frame)
            coef.est coef.se
(Intercept)  0.00     0.05
LRT          0.05     0.00
sch.avg     -0.01     0.01

Error terms:
 Groups   Name        Std.Dev. Corr
 school   (Intercept) 0.24
          LRT         0.01     0.62
 Residual             0.79
```



Random Intercept, Cross-Level Interaction

```
> lmer.6 <- lmer(Y ~ LRT * sch.avg +
+   (1 + LRT | school),
+   data=school.frame)
Model failed to converge with
max|grad| = 0.0309104 (tol = 0.002)
> display(lmer.6)
lmer(formula = Y ~ LRT * sch.avg +
(1 + LRT | school), data =
school.frame)
             coef.est coef.se
(Intercept) -0.01      0.05
LRT          0.05      0.00
sch.avg      0.00      0.01
LRT:sch.avg  0.00      0.00
Error terms:
 Groups   Name        Std.Dev. Corr
 school   (Intercept) 0.24
          LRT         0.01     0.64
 Residual             0.79
```



Which Model Fits Best (so far…)?

```
> AIC.ml <- function(M) {AIC(update(M,REML=F))}        ## We'll talk about
> BIC.ml <- function(M) {BIC(update(M,REML=F))}        ## AIC/BIC/DIC and
> DIC.ml <- function(M) {extractDIC(update(M,REML=F))} ## REML next week
> res <-
+ rbind(AIC=sapply(list(lmer.1,lmer.2,lmer.3,lmer.4,lmer.5,lmer.6),AIC.ml),
+       BIC=sapply(list(lmer.1,lmer.2,lmer.3,lmer.4,lmer.5,lmer.6),BIC.ml),
+       DIC=sapply(list(lmer.1,lmer.2,lmer.3,lmer.4,lmer.5,lmer.6),DIC.ml))
> colnames(res) <- c("lmer.1","lmer.2","lmer.3","lmer.4","lmer.5","lmer.6")
>
> t(round(res,2))
           AIC     BIC     DIC
lmer.1 5528.35 5545.12 5522.35  Y ~ 1 + (1 | school)
lmer.2 4752.30 4774.66 4744.30  Y ~ LRT + (1 | school)
lmer.3 4749.36 4782.90 4737.36  Y ~ LRT + (LRT | school)
lmer.4 4761.55 4795.08 4749.55  Y ~ LRT.cwc + (1 + LRT.cwc | school)
lmer.5 4751.02 4790.15 4737.02  Y ~ LRT + sch.avg + (1 + LRT | school)
lmer.6 4745.94 4790.65 4729.94  Y ~ LRT * sch.avg + (1 + LRT | school)
```

Some Automatic & Exact Methods

- There are a number of R packages that will do variable selection for lmer models, including:
  - `LMERConvenienceFunctions` automates backwards selection of fixed effects and forward selection of random effects, using AIC, BIC, etc.
    - `fitLMER.fnc()` is general-purpose function for this
  - `RLRsim` provides simulation-based exact likelihood ratio tests for random effects
    - `exactLRT()` performs exact LRT test for true ML fits
    - `exactRLRT()` performs exact LRT test for REML fits

Automated Variable Selection…

```
> library(LMERConvenienceFunctions) # for fitLMER.fnc() function...
# start with a "big fixed effects" model
> lmer.10 <- lmer(Y ~ LRT + VR + Gender + School.gender +School.denom +
+ (1+LRT|school), data=school.frame)
> lmer.11 <- fitLMER.fnc(lmer.10,
+ ran.effects=c("(School.gender|school)",
+ "(School.denom|school)"),method="BIC")
> anova(lmer.5,lmer.10,lmer.11)
refitting model(s) with ML (instead of REML)
Data: school.frame
Models:
lmer.11: Y ~ LRT + VR + Gender + (1 + LRT | school)
lmer.5: Y ~ LRT + School.denom + VR + (1 + LRT | school)
lmer.10: Y ~ LRT + VR + Gender + School.gender + School.denom + (1 + LRT |
lmer.10:     school)
        Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
lmer.11  9 4566.9 4617.2 -2274.4   4548.9
lmer.5  11 4577.2 4638.7 -2277.6   4555.2     0      2          1
lmer.10 14 4618.9 4697.2 -2295.5   4590.9     0      3          1
```

fitLMER.fnc:

1. Backwards elimination of F.E's
2. Forward selection of R.E.'s
3. Backwards elimination of F.E.'s

Exact Test of Random Effect..

```
library(RLRsim)

m0 <- lmer(Y ~ LRT + VR + Gender + (1  | school), data=school.frame)
lmer.11a <- lmer(Y ~ LRT + VR + Gender + (1|school) + (0 + LRT | school),
                 data=school.frame) # need indep rand effects for RLRsim...
lmer.LRT.only <- lmer(Y ~ LRT + VR + Gender + (0 + LRT | school),
                      data=school.frame)

formula(m0) # formula under H0: no random slopes for LRT
formula(lmer.11a) # model under HA: yes random slopes for LRT
formula(lmer.LRT.only) # model with *only* random slopes for LRT

exactRLRT(lmer.LRT.only,lmer.11a,m0)

#          simulated finite sample distribution of RLRT.
#
#          (p-value based on 10000 simulated values)
#
# data:
# RLRT = 6.2561, p-value = 0.0055
```

Summary

- Marginal and Conditional Models
- Estimation
  - MLE: Full maximum likelihood
  - EB: Empirical Bayes
  - REML: Restricted or Residual maximum likelihood
- Likelihood Ratio Tests, AIC, BIC
  - Change from REML to MLE for all three
- Df for marginal and conditional models
  - DIC, cAIC
- Variable selection: Practical Advice
- Example (London Schools)