Mixed Effects Models

Haley Hedlin, PhD
Quantitative Sciences Unit
Stanford Department of Medicine

April 29, 2015

HERC Cyberseminar

Outline

- What are you talking about?
- Why do I need mixed effects models?
- How do I fit mixed effects models?
- What can I conclude from mixed effects models
- What else should I know?

What are you talking about?

Alternative names for mixed effects models

- Mixed model, LMM/GLMM
- Random effects model, mixed effects model
- Longitudinal regression, repeated measures model
- Hierarchical model, multilevel model
- Covariance pattern model

Similar models

- GEE, marginal model, population average model
- Transition model

What are you talking about?

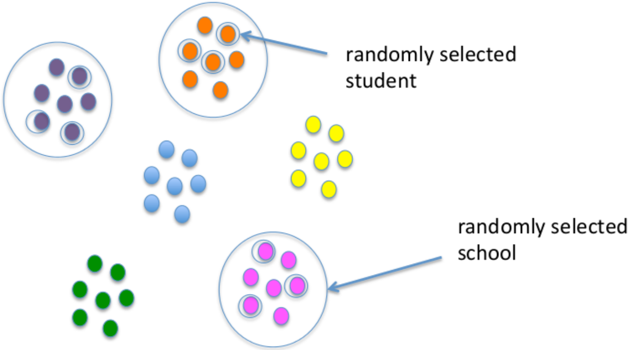
Often, people use the data type, design, or the software names as stand-ins for the model names.

Data and design types: clustered, longitudinal, spatial, pre-post, panel studies

Software

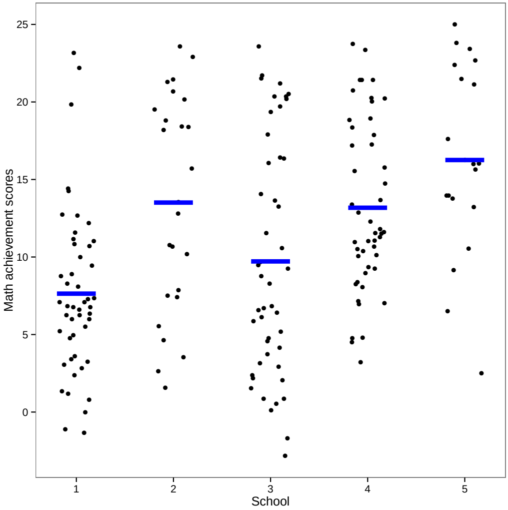
- R: lme in nlme package, lmer in lme4, glmm in glmm
- Stata: xtmixed, xtme, xtpoisson, gllamm
- SAS: proc mixed, proc glimmix

Example of clustered data



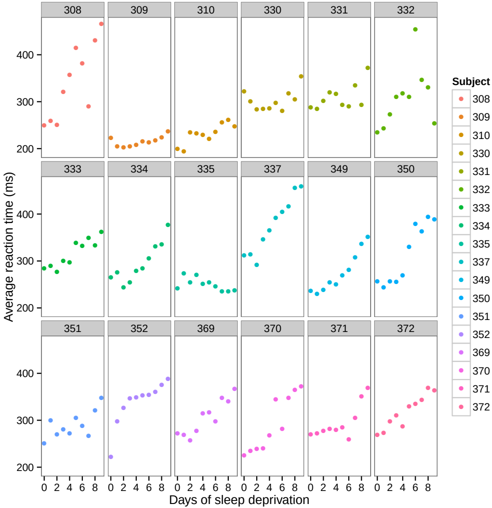
Example of clustered data

Record students' math achievement scores where students were randomly selected from schools in a state.



Example of longitudinal data

18 participants are in a study and are sleep deprived for several days. Each day, we measure that subject's reaction time.



What are you talking about?

Mixed models refers to the mixed effects in these models - random and fixed effects

Random effects have a distribution, fixed effects are constants

The random effects distribution has variance parameters that are estimated

What is the difference between longitudinal data and time series?

Univariate time series data typically arise from the collection of many data points over time from a single source, such as from a person, country, financial instrument, etc.

Longitudinal data typically arise from collecting a few observations over time from many sources, such as a few blood pressure measurements from many people.

There are some multivariate time series that blur this distinction but a rule of thumb for distinguishing between the two is that time series usually have more repeated observations than units/subjects while longitudinal data have more subjects than repeated observations.

What is the difference between longitudinal data and time series?

Longitudinal data analysis

- Repeated measures on a random sample
- Research questions about a population

Time series analysis

- Source(s) are rarely randomly sampled
- Rarely inference to a population

Correlation structure over time important in both

What is the relationship between longitudinal data and time series?

ARIMA (autoregressive integrated moving average) and ARMA models are often used to fit time-series data

Conditional autoregressive (CAR) and simultaneous autoregressive (SAR) models are commonly applied to spatial data

ARIMA/ARMA, CAR/SAR, and penalized splines (among many other models) can be parameterized as mixed linear models

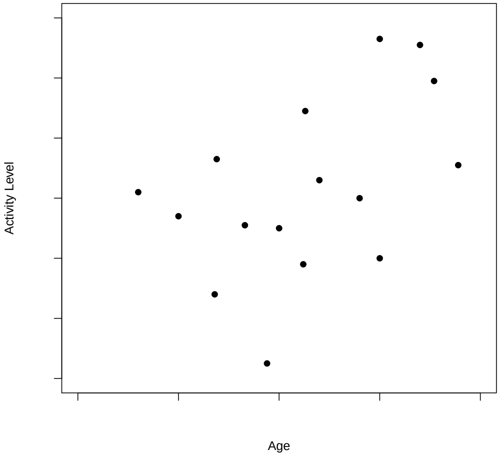
See:
Richly Parameterized Linear Models: Additive, Time Series, and Spatial Models Using Random Effects by James S. Hodges
and
New Introduction to Multiple Time Series Analysis by Helmut Lütkepohl

Outline

Why do I need these models?

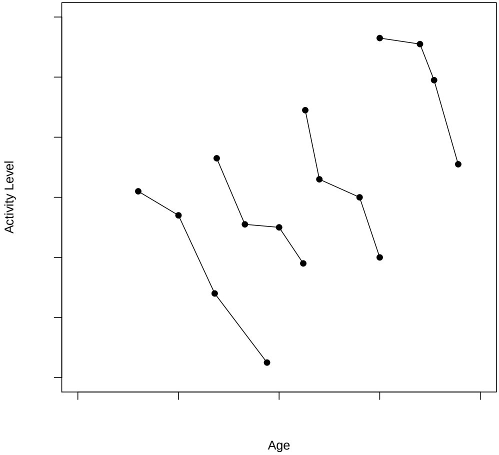
To avoid misrepresenting the direction of the effect.

Suppose we are treating longitudinal data as cross-sectional



Why do I need these models?

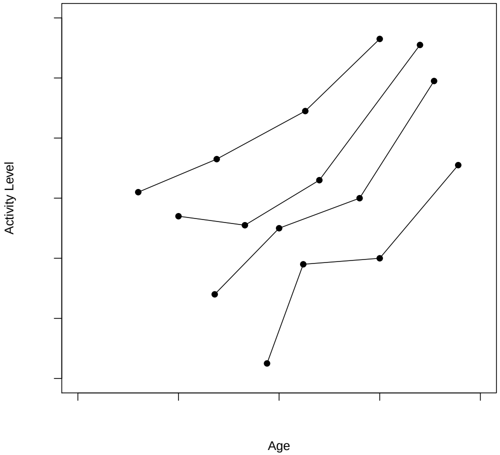
Now we correctly treat the data as repeated measures on individuals.



Why do I need these models?

To estimate cross-sectional and longitudinal effects.

We can use each subject as its own control.



Why do I need these models?

Your data was not collected with simple random sampling (SRS) and you cannot assume independence between all of your observations.

- Repeated measures, recurrent events, or longitudinal data
- Pre-post design
- Cluster or multistage sampling
- Spatial data
- Any combination of these data types

Why do I need these models?

If we ignore the correlations between observations, we will have

- inefficient estimation $\rightarrow$ loss of power
- estimated incorrect standard errors, confidence intervals, and p-values
- sub-optimal protection against biases caused by missing data
- lost the ability to model the correlation structure

Why do I need these models?

Think of a simple pre-post study where subjects have measurements before and after an intervention.

$$\mathrm{Var}(\bar{Y}_{\text{post}} - \bar{Y}_{\text{pre}}) = \mathrm{Var}(\bar{Y}_{\text{post}}) + \mathrm{Var}(\bar{Y}_{\text{pre}}) \text{ - } 2\mathrm{Cov}(\bar{Y}_{\text{post}}, \bar{Y}_{\text{pre}})$$

Ignoring the repeated measures on subjects will overestimate the variance when the covariance is positive.

Outline

How do I display longitudinal and clustered data?

Before fitting any model, you want to explore and look at your data.

Let's revisit the longitudinal and clustered datasets I introduced above.

For more on data displays – most textbooks on the topic devote an early chapter to the topic. I especially like Diggle, Heagerty, Liang, and Zeger's discussion on displaying longitudinal data.

How do I fit mixed models?

First, make sure your data is in the proper format.

Wide format:

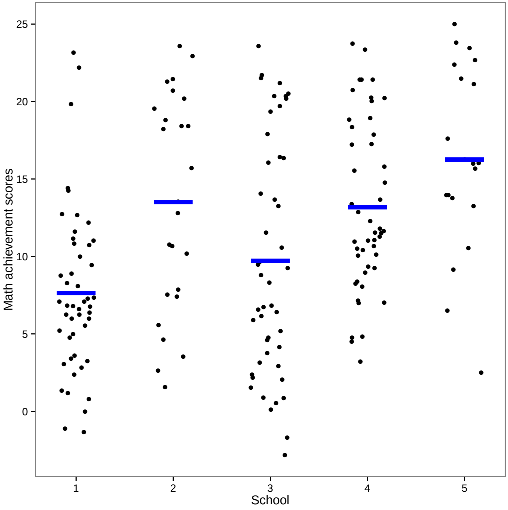
	Subject	Reaction.0	Reaction.1	Reaction.2	Reaction.3
1	308	249.5600	258.7047	250.8006	321.4398
2	309	222.7339	205.2658	202.9778	204.7070
2	310	199.0539	194.3322	234.3200	232.8416
4	330	321.5426	300.4002	283.8565	285.1330

How do I fit mixed models?

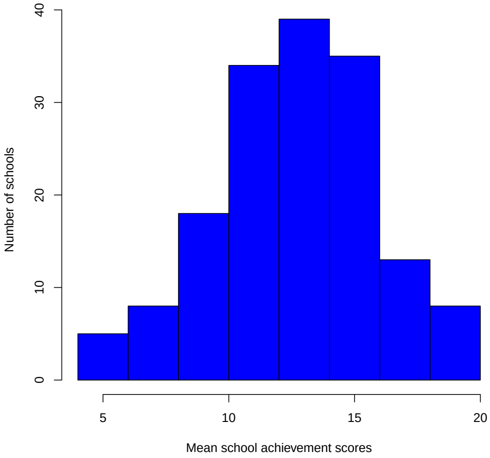
Long format:

```
  Reaction Days Subject
1 249.5600    0     308
2 258.7047    1     308
3 250.8006    2     308
4 321.4398    3     308
```

Clustered data



Clustered data



Two-stage normal normal model

$$
\begin{aligned}
Y_{ij} &= \theta_j + \epsilon_{ij} \\
\epsilon_{ij} &\sim N(0, \sigma_j^2) \\
\theta_j &\sim N(\theta, \tau^2)
\end{aligned}
$$

Y_{ij} is the score observed in school j from student i; the ϵ_{ij}'s are independent

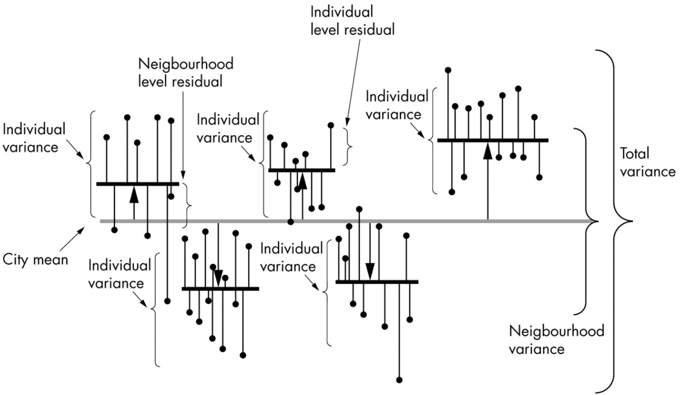
θ_j is the true average score for the school; the θ_j's are independent

σ_j^2 is the within school uncertainty

τ^2 is the heterogeneity across schools

θ is the true mean score

Two-stage normal normal model



Merlo, Juan, et al. "A brief conceptual tutorial of multilevel analysis in social epidemiology: linking the statistical concept of clustering to the idea of contextual phenomenon." Journal of epidemiology and community health 59.6 (2005): 443-449.

Intraclass correlation

$$
\begin{aligned}
Y_{ij} &= \theta_j + \epsilon_{ij} \\
\theta_j &= \theta + b_j, \text{ where } b_j \sim N(0, \tau^2) \\
Y_{ij} &= \theta + b_j + \epsilon_{ij} \\
Var(y_{ij}) &= Var(b_j) + Var(\epsilon_{ij}) = \tau^2 + \sigma_j^2
\end{aligned}
$$

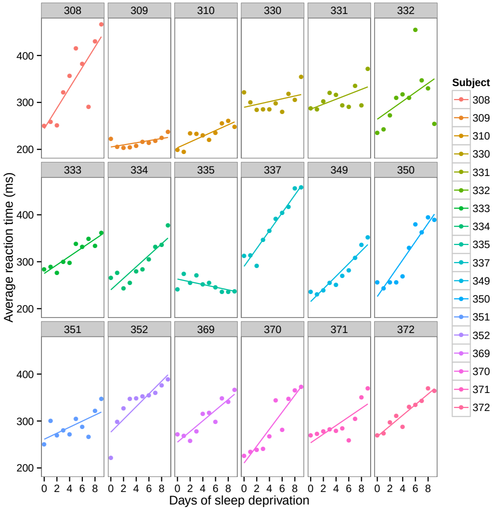
σ_j^2 is the within school variance for school j

Note: often we assume $\sigma_j^2 = \sigma^2$

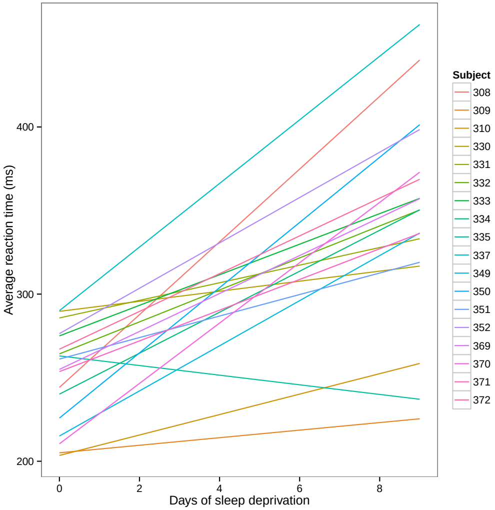
τ^2 is the between school variance

The intraclass correlation coefficient (ICC) is $\frac{\tau^2}{\tau^2+\sigma^2}$

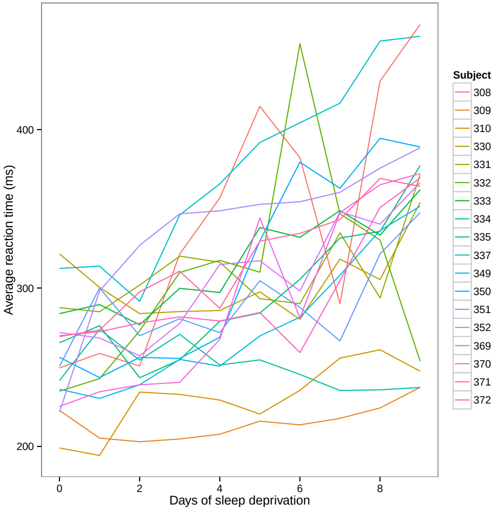
Plotting longitudinal data



Plotting of longitudinal data

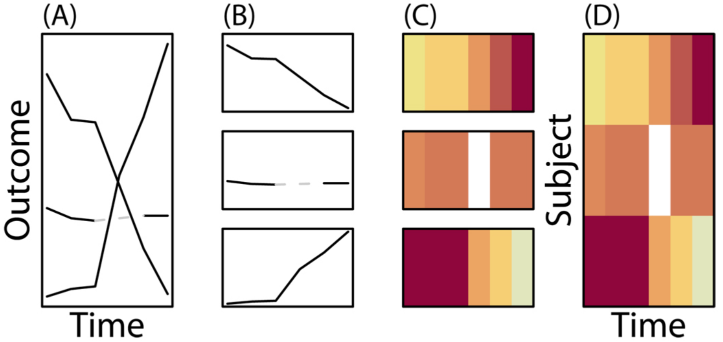


Plotting of longitudinal data



Plotting of longitudinal data

A (saucy) alternative to spaghetti plots for categorical data are lasagna plots.



Swihart, B. J., Caffo, B., James, B. D., Strand, M., Schwartz, B. S., & Punjabi, N. M. (2010). Lasagna plots: a saucy alternative to spaghetti plots. Epidemiology (Cambridge, Mass.), 21(5), 621.

Mixed effects model

Allow a random intercept, b_{0i}:

$$time_{ij} = \beta_0 + \beta_1 * days_{ij} + b_{0i} + \epsilon_{ij}$$

Add in a random slope term, b_{1i}:

$$time_{ij} = \beta_0 + \beta_1 * days_{ij} + b_{0i} + b_{1i} * days_{ij} + \epsilon_{ij}$$

where $\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_0^2 & \tau_{10} \\ \tau_{10} & \tau_1^2 \end{pmatrix}\right)$ are independent over i and with $\epsilon_{ij} \sim N(0, \sigma^2)$

Mixed effects model

In general

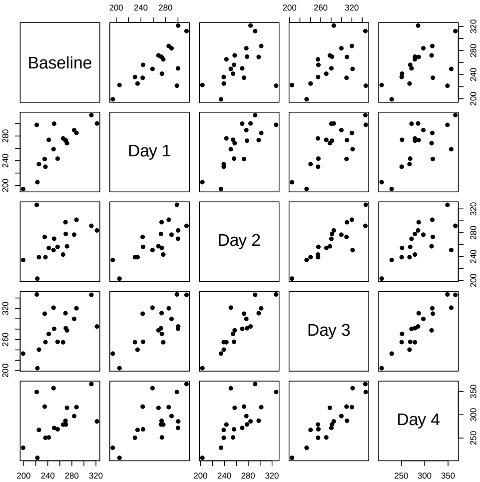
$$\begin{aligned}\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon} \\ \mathbf{b} &\sim N(\mathbf{0}, \mathbf{D}) \\ \boldsymbol{\epsilon} &\sim N(\mathbf{0}, \sigma^2 \mathbf{I})\end{aligned}$$

or, to allow a broader class of models,

$$\begin{aligned}\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon} \\ \mathbf{b} &\sim N(\mathbf{0}, \mathbf{G}(\phi_G)) \\ \boldsymbol{\epsilon} &\sim N(\mathbf{0}, \mathbf{R}(\phi_R))\end{aligned}$$

What's $\mathbf{D}$ or $\mathbf{G}(\phi_G)$?

Scatterplot matrix of reaction times



Empirical correlation matrix

	Day 0	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8
Day 0	1.00	0.74	0.47	0.46	0.45	0.37	0.22	0.49	0.33
Day 1	0.74	1.00	0.77	0.74	0.65	0.53	0.31	0.48	0.40
Day 2	0.47	0.77	1.00	0.88	0.69	0.49	0.45	0.59	0.41
Day 3	0.46	0.74	0.88	1.00	0.91	0.72	0.67	0.60	0.60
Day 4	0.45	0.65	0.69	0.91	1.00	0.85	0.75	0.69	0.74
Day 5	0.37	0.53	0.49	0.72	0.85	1.00	0.74	0.69	0.90
Day 6	0.22	0.31	0.45	0.67	0.75	0.74	1.00	0.70	0.73
Day 7	0.49	0.48	0.59	0.60	0.69	0.69	0.70	1.00	0.76
Day 8	0.33	0.40	0.41	0.60	0.74	0.90	0.73	0.76	1.00
Day 9	0.52	0.55	0.42	0.57	0.72	0.84	0.46	0.66	0.88

Covariance structures

In covariance pattern models for longitudinal data, a structure is imposed on the covariance.

- Independence - used in ordinary linear regression
- Exchangeable, uniform, compound symmetry
- Toeplitz - used when there is equal spacing among observations
  - Autoregressive - e.g. AR(1)
  - Banded
- Unstructured

Covariance structures

Bias vs. precision

- Imposing too little structure on the covariance results in many parameters being estimated in the covariance and less efficient parameter estimation for the mean.
- Imposing too much structure results in model misspecification that could result in incorrectly estimating β.

Fitzmaurice, Laird, and Ware provide a nice discussion of covariance pattern models.

How do I fit mixed models?

Maximum likelihood estimation

For a given covariance matrix, we have a general linear model and can estimate β using generalized least squares.

Now we can plug $\hat{\beta}(\phi)$ into the profile log likelihood for ϕ where ϕ represents the unknown variance parameters.

Maximize to obtain MLE for ϕ using numerical maximization. However, using $\hat{\beta}$ in the estimation gives us a biased estimate of the variance.

How do I fit mixed models?

MLEs for variances are biased downward using regular maximum likelihood estimation.

Example: $y_i \sim N(\mu, \sigma^2)$
ML estimate $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$, but $\mathrm{E}(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2$.

Instead, use the restricted maximum likelihood (REML) estimate
$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$

How do I fit mixed models?

R: lme in nlme package, lmer in lme4, glmm in glmm

- nlme vs lme4 - nlme has more flexibility in specifying the covariance structure, lme4 allows you to fit GLMs
- glmm uses MCMC in fitting

Stata: xtmixed, xtme, xtpoisson, gllamm

- gllamm is the most flexible, but can also be slow and takes the longest to learn

SAS: proc mixed, proc glimmix

How do I check the fit of mixed models?

The diagnostics for mixed models are similar to diagnostics for regular linear regression.

Create plots to check that the fitted mean is consistent with the observed data and check that the estimated covariance is consistent with the empirical variance.

Variograms - a graph displaying variation due to random effects, serial correlation, and measurement error in longitudinal data; useful in model selection

See Diggle, Heagerty, Liang, and Zeger

Outline

What can I conclude from mixed effects models?

Mixed models - observations are correlated because they are from the same subject and share the same underlying processes. Interpretation is <u>conditional</u> on the underlying processes

Marginal models (GEE) - observations are <u>marginally</u> correlated. The mean and covariance are modelled separately. Inference about the mean can sometimes be made even when the covariance is incorrectly specified

Transition models - observations are correlated because the past influences the presence. These models assume observations have Markov property.

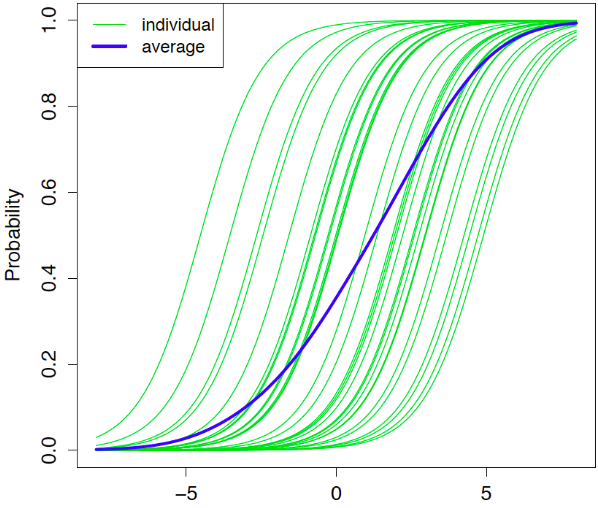
What can I conclude from mixed effects models?

In linear case, all three models can be formulated so the β coefficients have a marginal interpretation.

This is not true in the generalized case.

- In mixed models, the interpretation is conditional on the individual.
  - Subject-specific interpretation
- In marginal models, the interpretation is marginal.
  - Population averaged interpretation
- In transitional models, the interpretation is conditional on previous observations.

What can I conclude from mixed effects models?



What can I conclude from mixed effects models?

More topics –

Cross-level interactions

e.g. Does the effect of student SES on achievement depend on school-level SES? Perhaps effect is stronger in lower SES schools.

Contextual and compositional effects

e.g. Student outcomes are affected by the covariates of other students in the same school.

Level-2 endogeneity

e.g. SES is correlated with the school-level random intercepts. Some of the beneficial effects of unobserved school characteristics are falsely attributed to SES

Also, take special care with stochastic time-varying covariates

For more see Castellano, Katherine E., Sophia Rabe-Hesketh, and Anders Skrondal. "Composition, Context, and Endogeneity in School and Teacher Comparisons." Journal of Educational and Behavioral Statistics 39.5 (2014): 333-367.
and
Duncan, Craig, Kelvyn Jones, and Graham Moon. "Context, composition and heterogeneity: using multilevel models in health research." Social science & medicine 46.1 (1998): 97-117.

Outline

Empirical Bayes and shrinkage

Clusters with more precise mean estimates (due to larger number of observations or more precise measurements) contain more information than other clusters.

Information can be borrowed across clusters when making predictions.

Bayesian estimates (shrunken estimates) are commonly seen in health research ranking hospitals

See Hofer, T. P., Hayward, R. A., Greenfield, S., Wagner, E. H., Kaplan, S. H., & Manning, W. G. (1999). The unreliability of individual physician report cards for assessing the costs and quality of care of a chronic disease. Jama, 281(22), 2098-2105.
and
Merlo, Juan, et al. "A brief conceptual tutorial of multilevel analysis in social epidemiology: linking the statistical concept of clustering to the idea of contextual phenomenon." Journal of epidemiology and community health 59.6 (2005): 443-449.

What if I have missing data?

In completely random missingness (MCAR),

both GEE and mixed effects models give unbiased estimates.

In random missingness (MAR),

GEE gives biased estimates of mean

mixed effects models give unbiased estimates IF the mean and covariance structures are both correctly specified and multiple imputations methods used

In informative missingness (NMAR),

both GEE and mixed effects models give biased estimates.

Extensions and related topics

Meta-analysis

Joint survival analysis and longitudinal data

Tsiatis, Anastasios A., and Marie Davidian. "Joint modeling of longitudinal and time-to-event data: an overview." Statistica Sinica 14.3 (2004): 809-834.

Functional data analysis - treats data as curves

See Ramsay and Silverman (2002, 2005)

Sample size calculations can sometimes be derived but are usually done using simulations; be sure to account for within and between subject correlation

References

Diggle, P. J., Heagerty, P., Liang, K. Y., & Zeger, S. L. (2002). Analysis of longitudinal data. (2nd ed.). New York, NY: Oxford University Press

Fitzmaurice, Garrett M., Nan M. Laird, and James H. Ware (2012). Applied longitudinal analysis. Vol. 998. John Wiley & Sons.

Fitzmaurice, G. M., Davidian, M., Verbeke, G., & Molenberghs, G. (2008). Longitudinal data analysis. Boca Raton, FL: CRC Press.

Rabe-Hesketh, S., & Skrondal, A. (2012). Multilevel and longitudinal modeling using Stata. (3rd ed.) College Station, TX: Stata Press.

Contact

Haley Hedlin

hedlin@stanford.edu

http://med.stanford.edu/qsu.html

Extras

Pros and cons of mixed models

- $+$ Can easily accommodate unequally spaced time points and unbalanced data.
- $+$ Flexible, can easily incorporate additionally clustered data, smooth functions (penalized spline smoothing) etc.
- $\pm$ Parsimonious modeling of covariance. But random effects imply specific correlation structure, less flexibility.
- $+$ Allows likelihood-based inference.
- $-$ Fitting of models often hard (especially for generalized case).

Pros and cons of marginal models

- \+ Separate modeling of mean and correlation. Correlation model does not change interpretation of parameters.
- ± Are appropriate for inference about the population mean.
- \+ Only requires specification of the first two moments, not the entire likelihood (less assumptions).
- − No likelihood-based inference (but instead: GEE).
- \+ Can easily accommodate unequally spaced time points and unbalanced data.

Pros and cons of transition models

- − Most meaningful for equally spaced time points t_{ij} . What about unequally spaced data? Missing data?
- + Might be a very meaningful way to think of the underlying process in some cases (e.g. for categorical data when thinking of transition between "states")
- + Allows likelihood-based inference (typically conditional on first q observations).

```
# code to create figures and fit models for mixed effects models
# author: Haley Hedlin
# email: hedlin@stanford.edu
# date: April 29, 2015

# these packages need to be installed using install.packages()
# if you don't already have them installed on your machine
library(lme4)
library(nlme)
library(ggplot2)
library(lattice)
library(xtable) # to turn R output into latex tables


### Math Achievement Scores
# plot a few schools
sch.mn <- aggregate(MathAchieve$MathAch,
    by=list(School=MathAchieve$School), mean)
mathsub <- MathAchieve[MathAchieve$School %in% unique(MathAchieve$School)[1:5],]
mathsub$School <- as.numeric(mathsub$School)
sch.mn.sub <- aggregate(mathsub$MathAch,
    by=list(School=mathsub$School), mean)

pdf(file="scorepoints.pdf")
ggplot() + theme_bw() +
    theme(panel.grid.major = element_blank(),
    panel.grid.minor = element_blank()) +
    geom_jitter(data=mathsub, aes(School, MathAch), position =
 position_jitter(width = .2)) +
    geom_segment(aes(x = seq(0.8, 4.8, by=1), xend = seq(1.2, 5.2, by=1),
        y = sch.mn.sub$x, yend = sch.mn.sub$x), lwd=2, col="blue") +
    ylab("Math achievement scores")
dev.off()

pdf(file="SchScoreSistbn.pdf")
hist(sch.mn$x, breaks=10, xlab="Mean school achievement scores",
    main="", ylab="Number of schools", col="blue")
dev.off()

### sleep deprivation data
# xyplot in lattice

pdf(file="sleeplattice.pdf")
xyplot(Reaction ~ Days | Subject, sleepstudy, type = c("g","p","r"),
        index = function(x,y) coef(lm(y ~ x))[1],
        xlab = "Days of sleep deprivation",
        ylab = "Average reaction time (ms)", aspect = "xy")
dev.off()
```

```r
# same plot using faceting in ggplot
pdf(file="sleepggplotpts.pdf")
p <- ggplot(sleepstudy, aes(Days, Reaction, group=Subject,
            color=Subject)) +
        theme_bw() + xlab("Days of sleep deprivation") +
        ylab("Average reaction time (ms)") +
        theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank()) +
        scale_x_continuous(breaks=seq(0,8,by=2))
pfw <- p + geom_point() + facet_wrap(~Subject, nrow=3)
pfw
dev.off()

# add lines
pdf(file="sleepggplotlines.pdf")
pfw + stat_smooth(method=lm, se=FALSE)
dev.off()

pdf(file="sleepspaghetti.pdf")
p + geom_line()
dev.off()

pdf(file="sleeplines.pdf")
p + stat_smooth(method=lm, se=FALSE)
dev.off()


# to create a correlation matrix showing the correlations over time, need to put
 the data into wide format
head(sleepstudy)
?reshape
sleepwide <- reshape(sleepstudy, v.names="Reaction", timevar="Days",
                    idvar="Subject", direction="wide")
head(sleepwide)

# scatterplot matrix
pdf(file="sleepmatrix.pdf")
pairs(sleepwide[,c(2:6)], labels=c("Baseline", "Day 1", "Day 2",
        "Day 3", "Day 4"), pch=19)
dev.off()

# correlation matrix
cor(sleepwide[,-1])
round(cor(sleepwide[,-1]),2)
print(xtable(round(cor(sleepwide[,-1]),2)))
```

```r
# model fitting

(fm1 <- lmer(Reaction ~ Days + (Days|Subject), sleepstudy))
(fm2 <- lmer(Reaction ~ Days + (1|Subject) + (0+Days|Subject), sleepstudy))



### activity level vs. age - from Hayat & Hedlin

x <- c(3,5,6.8,9.4, 6.9,8.3,10,11.2, 11.3,12,14,15, 15,17,17.7,18.9,
       18,19.4,19.9,22.4, 21.7,21.9,24,25.8, 24.8,25.3,26.6,27.2)
y <- c(13.8,14.6,17.2,19.5, 12.7,14.9,15,16.2, 11.1,13.4,14,16, 8.7,8.9,10.1,12.9,
       5.1,6.5,8,9.1, 3.9,5.1,5.2,6.9, 3,3.9,4.1,6.5)*-1

pdf(file="misleading.pdf")
plot(x[1:16],y[1:16],xlab="Age",ylab="Activity Level",xlim=c(0,20),ylim=c(-20,-8),
     pch=19, axes=FALSE)
axis(2,at=seq(-20,-8,by=2),lab=rep("",7))
axis(1,at=seq(0,20,by=5),lab=rep("",5))
box()
dev.off()

# add the connecting lines
pdf(file="FullStory.pdf")
plot(x[1:16],y[1:16],xlab="Age",ylab="Activity Level",xlim=c(0,20),ylim=c(-20,-8),
     pch=19, axes=FALSE)
axis(2,at=seq(-20,-8,by=2),lab=rep("",7))
axis(1,at=seq(0,20,by=5),lab=rep("",5))
box()
for(i in 1:4){
  lines(x[((i-1)*4+1):(i*4)],y[((i-1)*4+1):(i*4)])
}
dev.off()

# add the connecting lines in the other direction
pdf(file="FullStoryUp.pdf")
plot(x[1:16],y[1:16],xlab="Age",ylab="Activity Level",xlim=c(0,20),ylim=c(-20,-8),
     pch=19, axes=FALSE)
axis(2,at=seq(-20,-8,by=2),lab=rep("",7))
axis(1,at=seq(0,20,by=5),lab=rep("",5))
box()
for(i in 1:4){
  lines(x[seq(i, i+12, by=4)],y[seq(i, i+12, by=4)])
}
dev.off()
```