Review of GEE and Mixed Models

Robin Mejia

Email: *rjmejia@berkeley.edu*

Office Hours: Thurs, 10 - 11am, 111 Haviland

October 28, 2015

GEE Recap

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + e_{ij}$$

GEE is an estimation method. We use GEE to estimate a marginal model.

- Is GEE always different from the linear models you've seen before?
- In what way can it be different?
- What is the interpretation of the coefficient on your explanatory variable?
- What assumptions do we make when we fit a GEE model?

GEE Recap

- Is GEE always different from the linear models you've seen before?
  - No. When you chose an independent correlation structure, GEE will give the same estimate as OLS.
- In what way can it be different?
  - GEE allows you to take advantage of the correlation structure to estimate the mean, which can give a more efficient estimate (smaller variance)
- What is the interpretation of the coefficient on your explanatory variable?
  - The population average effect. For linear models, this will be the same estimate as an individual effect, but for others, such as logistic regression, it will be different.
- What assumptions do we make when we fit a GEE model?
  - Very few. The model allows for inference that is robust to misspecification of the correlation structure.

GEE - mean model

GEE uses the correlation structure you specify to build a weight matrix used to estimate the mean.

- Independence: same as models you already know.
- Exchangeable: correlation of all measurements for an individual are the same (ρ)
  - Common choice. If you specify robust standard errors, it will give you SE's that don't assume correlation structure is correct.
- Autoregressive: correlation decays with distance in time
  - Good for a data such as height measurements every six months, etc. An exponential correlation matrix generalizes this for unequally spaced measurements.
- Unstructured: We don't assume anything.
  - A great idea and, but it requires a lot of data. Frequently, the model won't converge.
- These are commonly used options. There are others, and Stata allows you to specify your own.

GEE - Inference

- The estimation of standard errors does not have to use the correlation structure you supply for estimation of the mean model. In Stata, the default is to use the correlation structure you supply *unless* you request robust standard errors.
- The take away: if you choose robust standard errors, your standard errors will be okay even if the true correlation structure is different from what you specified.
- However, you need a decent amount of data to estimate robust standard errors.

Mixed Model Recap

- How is a model with random effects different the linear models you've seen before?
- In a mixed model, what is the interpretation of coefficient estimate on your explanatory variable?
- What assumptions do we make when we fit a random effects model?
- When does a mixed effects model have the same correlation structure as a GEE model?

Mixed Model Recap

- How is a model with random be different the linear models you've seen before?
  - A random effect is a latent variable; we estimate a variable we cannot observe in the data
- what is the interpretation of coefficient estimate on your explanatory variable?
  - An individual effect.
- What assumptions do we make when we fit a random effects model?
  - We assume we know the structure of the entire data generating distribution, including that we know the distribution of the unobserved latent variable(s) (the random effect(s))
- When does a mixed effects model have the same correlation structure as a GEE model?
  - A mixed model with a only a random intercept implies an exchangeable correlation structure

Mixed Model Recap

$$
\begin{aligned}
Y_{ij} &= \beta_0 + \beta_{0i} + \beta_1 x_{ij} + e_{ij} \\
Y_{ij} &= \beta_0 + \beta_{0i} + (\beta_1 + \beta_{1i}) x_{ij} + e_{ij}
\end{aligned}
$$

- What is are the random effects in this model?
- What do we usually assume about the distribution of the random effects?

Mixed Model Recap

- What does a mixed model give us that marginal model does not?
  - Partition of variance: estimating how much variance occurs within an individual (or independent unit) and how much variance is attributable to differences between units
  - A mixed model can give us individual effects, and can allow for those individual effects to vary
- What components of a model can be random?
  - A random intercept allows each individual to have a different baseline
  - A random slope allows a covariate to have a different effect on each individual

Hierarchical Models

$$Y_{ijk} = \beta_0 + \beta_{0i} + \beta_{0ij} + \beta_1 x_{ijk} + e_{ijk}$$

We can estimate random effects at multiple levels. For example:

- A school effect and a classroom effect within a school
- A village effect and a household effect
- A household effect and a person effect
- A person effect and a visit effect

A very abbreviated comparison of approaches

	Mixed Model	GEE
Focus	Variance components Regression coefficients	Regression coefficients
Parameter	Individual effect	Population average effect
Assumptions	Correct specification	Large enough sample

Example

You are analyzing a cluster-randomized village level intervention to improve water sanitation. Your outcome is whether or not a child experienced a major diarrhea event. All children under five are assessed for diarrheal events.

- What analysis should you conduct if you want to assess the effect of the intervention on the average diarrheal events per child in the entire community

Example

This question asks for a population average effect. As the outcome is binary, you would likely fit a logistic regression and use GEE to estimate, allowing for measurements to be correlated at the village level.

- What would you do if you want to know whether there is significant variation in diarrheal events among villages?