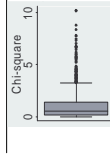


Repeated Measures, Part 3

Charles E. McCulloch,
Division of Biostatistics,
Dept of Epidemiology and Biostatistics,
UCSF

May, 2009



Outline

1. More on XTGEE
2. Examples
3. Robust standard errors
4. Binary outcomes
5. Changing the link function
6. Modeling practice
7. Summary

2



More on xtgee

Recall the command format:

```
xtgee depvar predvars,  
      family(distribution)  
      link(how to relate mean to predictors)  
      corr(correlation structure)  
      i(cluster variable)  
      t(time variable)  
      robust
```

3



More on xtgee

Here are the commonly used options:

Family

- binomial
- Gaussian (i.e., normal) default
- gamma
- nbinomial
- Poisson



4

More on xtgee

Link

- identity (model mean directly) default for Gaussian
- log default for Poisson
- logit default for binomial
- power
- probit



5

More on xtgee: main menu

Family and link choices:	Gaussian	Inverse Gaussian	Binomial	Poisson	Negative binomial	Gamma
Identity	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Log	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Logit	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Probit	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
C log-log	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Power	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Odds power	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Neg binom.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Reciprocal	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



More on xtgee

Correlation structure

independent
exchangeable default
ar #
unstructured



7

Examples

Let's try this on the birthweight data.

```
. xtgee bweight birthord initage, family(gaussian)
  link(identity) corr(exchangeable) i(momid)

. xtgee bweight birthord initage, fam(gau) link(i)
  corr(exch) i(momid)

. xtgee bweight birthord initage, i(momid)
```

all give the same output:



8

Examples

Iteration 1: tolerance = 7.180e-13

```
GEE population-averaged model
Group variable:      momid      Number of obs   =   1000
Link:                identity   Number of groups =    200
Family:              Gaussian   Obs per group: min =     5
Correlation:         exchangeable max =     5
Scale parameter:    324458.3    Wald chi2(2)    =   30.87
                                      Prob > chi2     =   0.0000
```

bweight	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
birthord	46.608	9.944792	4.687	0.000	27.11657	66.09943
initage	26.73226	8.957553	2.984	0.003	9.175783	44.28874
_cons	2526.622	162.544	15.544	0.000	2208.042	2845.203



9

Robust standard errors

The "robust" option asks Stata to estimate the standard errors empirically from the data. This has the significant advantage that it gives valid standard errors *even when the assumed correlation structure is wrong*. It is also better than assuming an unstructured variance-covariance structure, because it bypasses the estimation of the correlations over "time" to directly get an estimate of the standard errors.

The robust option works well when there are many "subjects" and not too much data per subject and not much missing data. So, for example, it would work very well when there are 2,000 subjects most measured yearly for four years. It would not work well if the "subjects" were 8 centers in a multi-center trial, each with 1,000 patients enrolled.



Examples: variation III (xtmixed)

```
xtmixed bweight birthord initage || momid:
Mixed-effects REML regression      Number of obs   =   1000
Group variable: momid              Number of groups =    200
                                   Obs per group: min =     5
                                   avg   =    5.0
                                   max   =     5

Log restricted-likelihood = -7649.3763      Wald chi2(2)     =    30.75
                                   Prob > chi2       =    0.0000
```

bweight	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
birthord	46.608	9.951014	4.68	0.000	27.10437 66.11163
initage	26.73226	9.002678	2.97	0.003	9.08734 44.37719
_cons	2526.622	163.3387	15.47	0.000	2206.484 2846.76

Random-effects Parameters		Estimate	Std. Err.	[95% Conf. Interval]
momid: Identity				
sd(_cons)		358.1759	23.71804	314.5797 407.8139
sd(Residual)		445.0229	11.13253	423.7298 467.386

LR test vs. linear regression: chibar2(01) = 209.20 Prob >= chibar2 = 0.0000



Binary outcomes/logistic regression

- Now let's take a look at the use of xtgee for clustered logistic regression. I took the Georgia babies data set and artificially dichotomized it as to whether birthweight was above or below 3000 grams.

- What options do we use now?

family =
link =
corr =



Binary outcomes: low birthweight

```
xtgee lowbirth birthord initage, i(momid) family(binomial)
```

```
GEE population-averaged model
Group variable:      momid      Number of obs   =   1000
Link:                logit      Number of groups =    200
Family:              binomial   Obs per group:  min =     5
Correlation:         exchangeable max =     5.0
Scale parameter:    1          Wald chi2(2)     =   11.30
                          Prob > chi2      =   0.0035
```

lowbirth	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
birthord	-.0829363	.0390214	-2.125	0.034	-.159417 - .0064557
initage	-.089028	.0337755	-2.636	0.008	-.1552267 - .0228293
_cons	1.267884	.6036077	2.101	0.036	.0848346 2.450933

How does this compare to the continuous data fit?
OK to assume exchangeable?

16

Binary outcomes: robust option

```
xtgee lowbirth birthord initage, i(momid) family(bino) robust
```

```
GEE population-averaged model
Group variable:      momid      Number of obs   =   1000
Link:                logit      Number of groups =    200
Family:              binomial   Obs per group:  min =     5
Correlation:         exchangeable max =     5
Scale parameter:    1          Wald chi2(2)     =   10.64
                          Prob > chi2      =   0.0049
```

(Std. Err. adjusted for clustering on momid)

lowbirth	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
birthord	-.0829363	.0384829	-2.16	0.031	-.1583614 - .0075113
initage	-.089028	.0341776	-2.60	0.009	-.1560149 - .0220412
_cons	1.267884	.6099252	2.08	0.038	.0724524 2.463315

17

Changing the link function

What kind of model would this command fit?

```
xtgee bweight birthord initage, i(momid) link(log)
```

```
GEE population-averaged model
Group variable:      momid      Number of obs   =   1000
Link:                log        Number of groups =    200
Family:              Gaussian   Obs per group:  min =     5
Correlation:         exchangeable max =     5
Scale parameter:    324595.5    Wald chi2(2)     =   30.56
                          Prob > chi2      =   0.0000
```

bweight	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
birthord	.0147553	.0031742	4.648	0.000	.0085339 .0209767
initage	.008179	.0027336	2.992	0.003	.0028212 .0135368
_cons	7.862211	.0503139	156.263	0.000	7.763598 7.960825

Interpretations:

18

Changing the link function

The log link implies we are modeling the log of the mean birthweight. The distribution has defaulted to normal. So our model is

$$\ln(\text{mean birthweight}) = \beta_0 + \beta_1 \text{birthord} + \beta_2 \text{initage}$$

or

$$\text{mean birthweight} = \exp(\beta_0 + \beta_1 \text{birthord} + \beta_2 \text{initage})$$

So with each increase of birth order by one the mean birthweight is predicted to change by $\exp(\beta_1) \times 100\%$, or in our case by $\exp(0.0148) = 1.015 \times 100\% = 101.5\%$ or an increase of 1.5%.



19

Changing the link function

What kind of model would this command fit?

```
xtgee lowbirth birthord initage,  
i(momid) family(bino) link(log)
```



20

Modeling practice 1

Epileptics were randomly allocated to a placebo or an anti-seizure drug (Progabide) group. The number of seizures was recorded during a baseline period and for four periods after beginning treatment.

Is the drug effective at reducing the number of seizures?

- family =
- link =
- corr =
- predictors =



21

Modeling practice 2

A quality improvement program was designed to reduce prescription of antibiotics for antibiotic resistant infections in emergency departments. 16 hospitals were randomized to the program or control group. Is the program effective?

family =
link =
corr =
predictors =

22



Notes on xtgee

- `xtgee` is a flexible regression command
- Handles a single level of clustering
- Handles a wide variety of distributions, links and correlation structures
- Five questions: distribution, predictors, link, correlation structure, cluster variable
- Not designed for inferences about the correlation structure
- Doesn't give predicted values for each cluster

23



Summary

- Hierarchical data structures are common.
- Lead to correlated data.
- Ignoring the correlation can be a serious error.
- `xtgee` can handle single level of clustering and a variety of outcome types.
- `xtmixed` can handle multiple levels for normally distributed data.
- Not discussed in class: `xtmelogit` and `xtmepoisson` can handle two levels of clustering for binary and Poisson outcomes. Random effects models and robust SEs are also available for time-to-event data.
- GEE methods have the advantage of robust SEs.
- Random effects models have the advantage of being able to generate predicted values and partition variability.

24