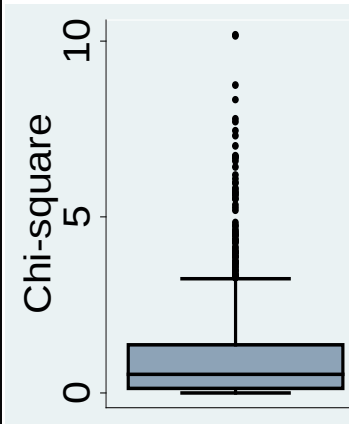


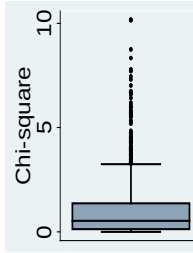
Repeated Measures, Lecture 3, Part 1

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

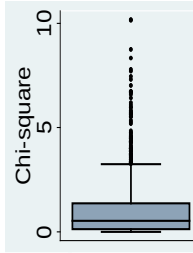


Outline

1. More on **mixed** and **xtgee**
2. Examples
3. Robust standard errors
4. Summary



MIXED for continuous outcomes

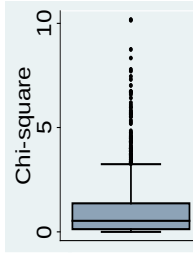


There can be multiple random effects specified by adding additional ||'s and random effects at the end. In this way you can handle multiple hierarchies or levels of clustering. Order is from highest to lowest level of clustering.

Example: backpain data

```
mixed logcost i.pracstyl i.educ ||  
  doctor: || patient:, reml
```

MIXED for continuous outcomes

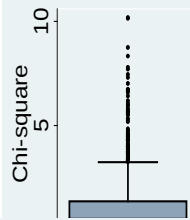


Specifying a random factor with a colon tells MIXED to allow different intercepts for each level of the factor, e.g., different intercepts for each patient.

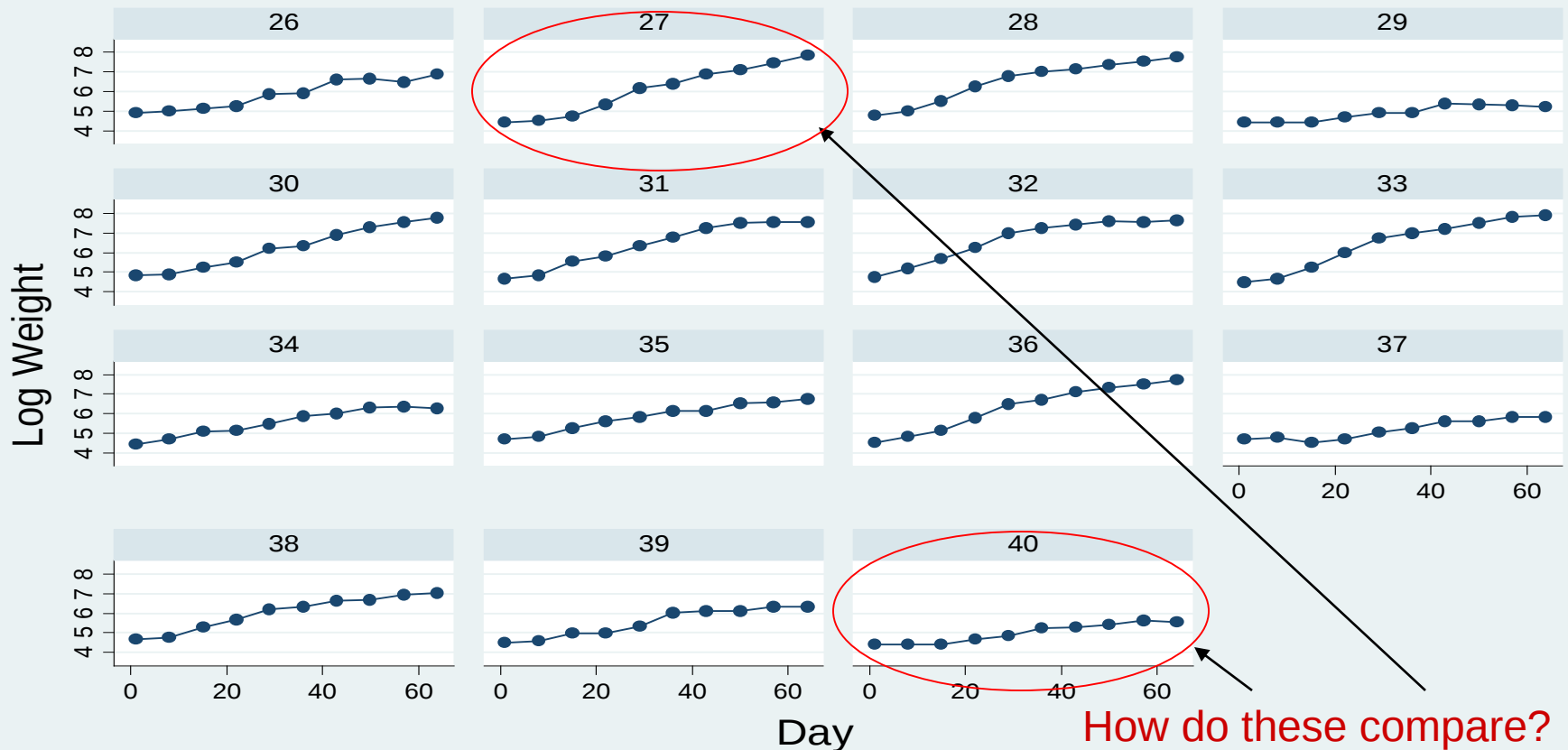
This is often sufficient, but there are cases in which more features of the model need to be included, for example, patient, animal or physician specific terms.

If so, these are added after the colon.

Mouse tumor/weight data

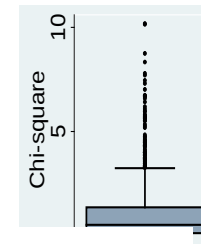


Mouse-specific trajectories of $\log(\text{weight})$
For a single treatment group



Graphs by mouse

Mouse tumor/weight data



- So - an adequate model would need animal specific slopes and intercepts.

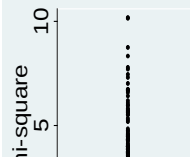
Stata code

- `mixed logw i.group day group#c.day
|| mouseid: day, cov(uns) reml`

Mouse specific
intercepts

Do the treatment groups
change differently over time
(assuming linearity)?
Mouse specific day
coefficients (slopes
with day).

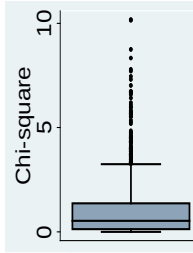
Mouse tumor/weight data



More details in lab 2, but the “|| mouse:” portion specifies that mouse is a random effect (i.e., allows a mouse specific intercept or constant term) and the “day” term specifies that the slopes over days that are associated with a mouse are also random effects. That is, it allows each mouse to have its own growth trend.

The “cov(un)” specifies that the variance and covariances of the random intercepts and slopes are unstructured. That is, the variances of the random intercepts and slopes are not assumed the same and they may be correlated.

Using xtgee: 5 things to specify



1. What is the distributional *family* (for fixed values of the covariates) that is appropriate to use for the outcome data? Normal, binary, binomial (other?).
2. Which predictors are we going to include in the model? (Nothing new here).
3. In what way are we going to *link* the predictors to the data? (Through the mean? Through the logit of the risk?)
4. What *correlation* structure will be used or assumed temporarily in order to form the estimates?
5. Which variable indicates how the data are clustered?

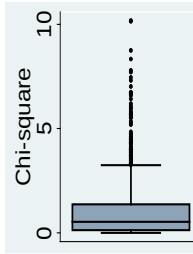
xtgee: syntax

Example: Fecal Fat

```
xtgee fecfat i.pilltype i.sex, i(patid)  
      t(pilltype) family(gaussian) corr(exch)
```

which can be shortened using defaults to

```
xtgee fecfat i.pilltype i.sex, i(patid)
```





More on xtgee

Here are the commonly used options:

Family

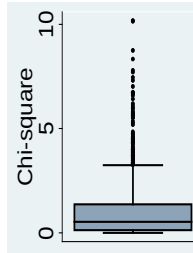
binomial

Gaussian (i.e., normal) default

gamma

nbinomial

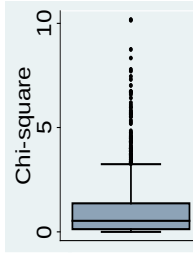
Poisson



More on xtgee

Link

identity (model mean directly)	<u>default for Gaussian</u>
log	<u>default for Poisson</u>
logit	<u>default for binomial</u>
power	
probit	



More on xtgee: main menu

xtgee - Fit population-averaged panel-data models by using GEE

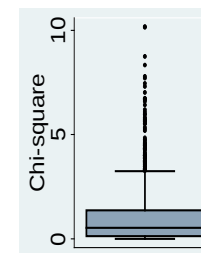
Model | Model 2 | Correlation | by/if/in | Weights | SE/Robust | Reporting | Optimization

Dependent variable: Independent variables: Panel settings...

Family and link choices:

	Gaussian	Inverse Gaussian	Binomial	Poisson	Negative binomial	Gamma
Identity	☒	☐	☐	☐	☐	☐
Log	☐	☐	☐	☐	☐	☐
Logit			☐			
Probit			☐			
C. log-log			☐			
Power	☐	☐	☐	☐	☐	☐
Odds power			☐			
Neg. binom.					☐	
Reciprocal	☐		☐	☐		☐

OK Cancel Submit



More on xtgee

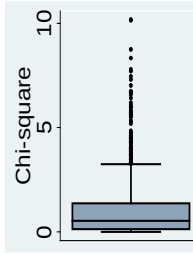
Correlation structure

independent

exchangeable default

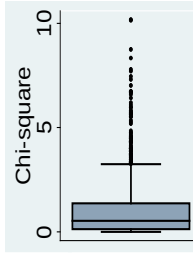
ar #

unstructured



Examples

Let's try this on the birthweight data.



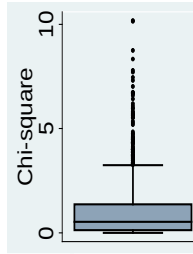
```
. xtgee bweight birthord initage, family(gaussian)
  link(identity) corr(exchangeable) i(momid)

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

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

all give the same output:

Examples



Iteration 1: tolerance = 7.180e-13

GEE population-averaged model		Number of obs	=	1000
Group variable:	momid	Number of groups	=	200
Link:	identity	Obs per group: min	=	5
Family:	Gaussian	avg	=	5.0
Correlation:	exchangeable	max	=	5
		Wald chi2(2)	=	30.87
Scale parameter:	324458.3	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

Examples

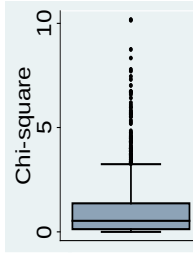
The command

```
. xtcorr
```

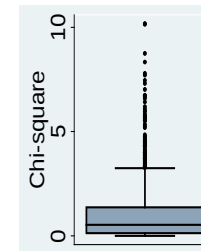
Estimated within-momid correlation matrix R:

	c1	c2	c3	c4	c5
r1	1.0000				
r2	0.3904	1.0000			
r3	0.3904	0.3904	1.0000		
r4	0.3904	0.3904	0.3904	1.0000	
r5	0.3904	0.3904	0.3904	0.3904	1.0000

gives the estimated correlation structure.



Examples: variation I



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

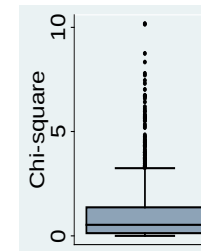
GEE population-averaged model		Number of obs	=	1000
Group and time vars:	momid birthord	Number of groups	=	200
Link:	identity	Obs per group: min	=	5
Family:	Gaussian	avg	=	5.0
Correlation:	unstructured	max	=	5
Scale parameter:	324495.1	Wald chi2(2)	=	30.43
		Prob > chi2	=	0.0000

bweight	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
birthord	44.70366	9.935604	4.499	0.000	25.23023 64.17708
initage	28.07164	8.79559	3.192	0.001	10.83261 45.31068
_cons	2505.539	159.0359	15.755	0.000	2193.834 2817.243

```
. xtcorr
Estimated within-momid correlation matrix R:
      c1      c2      c3      c4      c5
r1  1.0000
r2  0.1905  1.0000
r3  0.2977  0.4714  1.0000
r4  0.2475  0.4326  0.6417  1.0000
r5  0.3641  0.3814  0.4286  0.4487  1.0000
```

The estimated correlations are a bit different, but the coefficients and SEs have remained almost unchanged

Examples: variation II



```
. xtgee bweight birthord initage, i(momid) corr(uns) robust
```

```
Iteration 1: tolerance = .04763573
Iteration 2: tolerance = .00062083
Iteration 3: tolerance = .00001004
Iteration 4: tolerance = 1.668e-07
```

```
GEE population-averaged model
Group and time vars:      momid birthord
Link:                     identity
Family:                   Gaussian
Correlation:              unstructured

Number of obs      =      1000
Number of groups   =       200
Obs per group: min =         5
                  avg =        5.0
                  max =         5
Wald chi2(2)       =       29.05
Prob > chi2        =       0.0000

Scale parameter:      324495.1
```

(standard errors adjusted for clustering on momid)

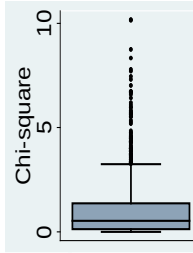
bweight	Coef.	Semi-robust Std. Err.	z	P> z	[95% Conf. Interval]	
birthord	44.70366	9.84746	4.540	0.000	25.40299	64.00432
initage	28.07164	9.141285	3.071	0.002	10.15506	45.98823
_cons	2505.539	161.11	15.552	0.000	2189.769	2821.308

Robust doesn't change
the coefficients, but
recalculates the SEs

Robust standard errors

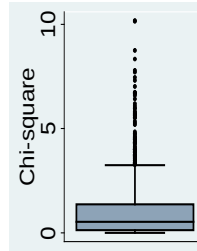
The “robust” option asks Stata to estimate the standard errors empirically from the data. This has the significant advantage that it gives valid standard errors *even when the assumed correlation structure is wrong*. It is also better than assuming an unstructured variance-covariance structure, because 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 (mixed)



```
mixed bweight birthord initage || momid:, reml
```

Mixed-effects REML regression
Group variable: momid

Number of obs = 1000
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

Summary

- `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
- `mixed` can handle multiple levels for normally distributed data.
- GEE methods have the advantage of robust SEs.
- Mixed effects models have the advantage of being able to handle more general types of designs and partition variability.

