GEE compared to Mixed modeling for LINEAR LINK

Pancreatic Enzyme example revisited:

```
proc mixed data = long1;
class pilltype personid;
model fat = pilltype/solution;
random intercept / subject = personid;
estimate "all compared to none" pilltype 1 1 -3 1;
run;
```

```
The Mixed Procedure

  Covariance Parameter Estimates
Cov Parm      Subject     Estimate
Intercept     personid      252.67        ←recall intraclass corr = > 252.67/(107+252.67) = 0.7025
Residual                    107.00

                        Solution for Fixed Effects
             NAME OF
             FORMER                    Standard
Effect       VARIABLE    Estimate       Error      DF    t Value    Pr > |t|
Intercept                 16.5333      7.7424       5       2.14      0.0858
pilltype     capsule       0.8833      5.9721      15       0.15      0.8844
pilltype     coated       14.5333      5.9721      15       2.43      0.0279
pilltype     none         21.5500      5.9721      15       3.61      0.0026
pilltype     tablet             0           .       .          .         .

        Type 3 Tests of Fixed Effects
              Num     Den
Effect         DF      DF    F Value    Pr > F
pilltype        3      15       6.26    0.0057

                                 Estimates
                                    Standard
Label                    Estimate       Error      DF    t Value    Pr > |t|
all compared to none    -49.2333     14.6287      15      -3.37      0.0042
```

```
proc genmod data = long1;
class pilltype personid;
model fat = pilltype/ dist = normal link = identity type3;
repeated subject = personid/type = exch corrw modelse;
estimate "all compared to none" pilltype 1 1 -3 1;
run;
```

```
The GENMOD Procedure

                Analysis Of GEE Parameter Estimates
                 Empirical Standard Error Estimates
                           Standard   95% Confidence
Parameter         Estimate    Error       Limits            Z Pr > |Z|

Intercept          16.5333   4.9644    6.8033  26.2634    3.33    0.0009
pilltype  capsule   0.8833   1.3991   -1.8588   3.6255    0.63    0.5278
pilltype  coated   14.5333   6.0039    2.7660  26.3007    2.42    0.0155
pilltype  none     21.5500   6.3279    9.1476  33.9524    3.41    0.0007
pilltype  tablet    0.0000   0.0000    0.0000   0.0000     .       .

                Analysis Of GEE Parameter Estimates
                Model-Based Standard Error Estimates
                           Standard   95% Confidence
Parameter         Estimate    Error       Limits            Z Pr > |Z|
Intercept          16.5333   7.7424    1.3585  31.7082    2.14    0.0327
pilltype  capsule   0.8833   6.3977  -11.6559  13.4225    0.14    0.8902
pilltype  coated   14.5333   6.3977    1.9941  27.0725    2.27    0.0231
pilltype  none     21.5500   6.3977    9.0108  34.0892    3.37    0.0008
pilltype  tablet    0.0000   0.0000    0.0000   0.0000     .       .
Scale              18.9649        .         .        .     .       .
NOTE: The scale parameter for GEE estimation was computed as the square root of the normalized
      Pearson's chi-square.


  Score Statistics For Type 3 GEE Analysis
                          Chi-
Source           DF      Square    Pr > ChiSq
pilltype          3        4.21        0.2402
```

Contrast Estimate Results

Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits	
all compared to none	-49.2333	-80.4471	-18.0195	-49.2333	15.9257	0.05	-80.4471	-18.0195

Contrast Estimate Results

Label	Chi-Square	Pr > ChiSq
all compared to none	9.56	0.0020

Working Correlation Matrix

	Col1	Col2	Col3	Col4
Row1	1.0000	0.6586	0.6586	0.6586
Row2	0.6586	1.0000	0.6586	0.6586
Row3	0.6586	0.6586	1.0000	0.6586
Row4	0.6586	0.6586	0.6586	1.0000

Exchangeable Working Correlation

Correlation	0.6585999815

SUMMARY:

1. Coefficient estimates are the same. (Marginal and conditional interpretation are the same for linear link). Standard errors are somewhat different since they are using different theory to represent uncertainty.

2. Notice the model based and empirical standard errors are not the same. With small number of clusters (here N=6) better to use model based standard errors for GEE (default is to use empirical "sandwich formula" standard errors, but when number of clusters is small this is difficult to estimate). As number of clusters increases, if working correlation matrix is correct and mean model is correct, then empirical standard error go to the model based standard errors asymptotically.

3. The Score Test which is used for overall effects based on GEE is not as powerful as the F-test from Proc Mixed (especially with smaller number of clusters).

4. Notice that in the GEE results we get the correlation estimate within cluster (i.e. intraclass correlation) whereas from Proc Mixed we get the partitioned within and between variances that can then be used to calculate the intraclass correlation. Not having the variance partitioned means that using GEE if we next added in Gender we would **not be able to calculate** a "% of person to person variability explained by gender".

GEE compared to Mixed modeling for LINEAR LINK

Math Achievement example revisited:

```
proc mixed data = math covtest;
class school;
model mathach = cses sector meanses cses*meanses cses*sector/ solution  ddfm = bw;
random intercept cses / subject = school type = un gcorr; run;
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	SCHOOL	2.3817	0.3717	6.41	<.0001
UN(2,1)	SCHOOL	0.1926	0.2045	0.94	0.3464
UN(2,2)	SCHOOL	0.1014	0.2138	0.47	0.3177
Residual		36.7212	0.6261	58.65	<.0001

Estimated G Correlation Matrix

Row	Effect	SCHOOL	Col1	Col2
1	Intercept	1224	1.0000	0.3919
2	cses	1224	0.3919	1.0000

The Mixed Procedure

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	12.1136	0.1988	157	60.93	<.0001
cses	2.9388	0.1551	7022	18.95	<.0001
SECTOR	1.2167	0.3064	157	3.97	0.0001
MEANSES	5.3391	0.3693	157	14.46	<.0001
cses*MEANSES	1.0389	0.2989	7022	3.48	0.0005
cses*SECTOR	-1.6426	0.2398	7022	-6.85	<.0001

```
proc genmod data = math;
class school;
model mathach = cses sector meanses cses*meanses cses*sector;
repeated sub = school/type = cs modelse; run;
```

GEE Model Information

Correlation Structure	Exchangeable
Subject Effect	SCHOOL (160 levels)
Number of Clusters	160
Correlation Matrix Dimension	67
Maximum Cluster Size	67
Minimum Cluster Size	14

Exchangeable Working Correlation

Correlation	0.0558478994

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	12.1139	0.1734	11.7740	12.4538	69.85	<.0001
cses	2.9358	0.1476	2.6466	3.2251	19.89	<.0001
SECTOR	1.2155	0.3080	0.6118	1.8191	3.95	<.0001
MEANSES	5.3392	0.3347	4.6831	5.9952	15.95	<.0001
cses*MEANSES	1.0441	0.3329	0.3916	1.6966	3.14	0.0017
cses*SECTOR	-1.6421	0.2372	-2.1070	-1.1771	-6.92	<.0001

Analysis Of GEE Parameter Estimates
Model-Based Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	12.1139	0.1928	11.7360	12.4917	62.84	<.0001
cses	2.9358	0.1509	2.6401	3.2316	19.46	<.0001
SECTOR	1.2155	0.2968	0.6338	1.7971	4.10	<.0001
MEANSES	5.3392	0.3578	4.6380	6.0404	14.92	<.0001
cses*MEANSES	1.0441	0.2914	0.4729	1.6152	3.58	0.0003
cses*SECTOR	-1.6421	0.2334	-2.0995	-1.1846	-7.04	<.0001
Scale	6.2481	.	.	.	.	.

NOTE: The scale parameter for GEE estimation was computed as the square root of the normalized Pearson's chi-square.

Using STATA for GEE:

```
. xi: xtreg mathach cses sector meanses  cses_meanses cses_sector, pa i(school)

OR

. xi: xtgee  mathach cses sector meanses  cses_meanses cses_sector, family(gaussian)
link(id) corr(exchangeable) i(school)


NOTE: xtreg, pa (the pa stands for 'population average') is equivalent to
xtgee, family(gaussian) link(id) corr(exchangeable)

By default it gives model based standard error estimates.  To get empirical standard
errors use the "robust" option.

Iteration 1: tolerance = .02735516
Iteration 2: tolerance = .00005525
Iteration 3: tolerance = 1.454e-07

GEE population-averaged model                   Number of obs      =      7185
Group variable:                       school    Number of groups   =       160
Link:                               identity    Obs per group: min =        14
Family:                             Gaussian                   avg =      44.9
Correlation:                   exchangeable                    max =        67
                                                Wald chi2(5)       =    783.93
Scale parameter:                    39.00605    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     mathach |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cses |     2.93584   .1508277    19.46   0.000     2.640223    3.231457
      sector |    1.215478   .2967278     4.10   0.000     .6339019    1.797053
     meanses |    5.339197   .3577137    14.93   0.000     4.638091    6.040303
cses_meanses |    1.044084   .2912786     3.58   0.000     .4731884    1.614979
 cses_sector |   -1.642067   .2332859    -7.04   0.000    -2.099299   -1.184835
       _cons |    12.11387   .1927583    62.84   0.000     11.73607    12.49167
------------------------------------------------------------------------------
```

Here are the resulting estimates if we completely ignore the clustering of kids within schools. In other words we treat the n=7185 kids as iid.

```
proc genmod data = math;
class school;
model mathach = cses sector meanses cses*meanses cses*sector;
run;
```

Analysis Of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	12.1024	0.1069	11.8929	12.3118	12827.1	<.0001
cses	1	2.9358	0.1552	2.6316	3.2400	357.73	<.0001
SECTOR	1	1.2706	0.1578	0.9613	1.5798	64.84	<.0001
MEANSES	1	5.1701	0.1908	4.7962	5.5440	734.52	<.0001
cses*MEANSES	1	1.0440	0.2998	0.4564	1.6315	12.13	0.0005
cses*SECTOR	1	-1.6420	0.2401	-2.1125	-1.1714	46.78	<.0001
Scale	1	6.2451	0.0521	6.1438	6.3481		

SUMMARY:

1. Coefficient estimates are nearly the same between random effects model and using GEE and also nearly the same as model that completely ignores correlation (although standard errors are diff for iid model see point 5 below).

2. Standard errors are very similar between RE and GEE. Notice that now the empirical and model based SE for GEE are more similar this is because we have larger number of clusters.

3. From the GEE estimation we find the intraclass correlation (after controlling for all individual and school level covariates) to be .0558. This is similar to the what we would get from the Mixed model variance components: 2.3817/(2.3817 + .101 + 2*.1926 + 36.72) = 0.0602.

4. As mentioned in the previous example, from the GEE we do not get any types of estimates about variability of slopes or variability of intercepts. Nevertheless if all you care about is inference for the fixed effects (both individual and school level covariates and their interactions) then you get what you need from GEE.

5. Notice that for SECTOR and MEANSES in the iid model that the standard errors are **too small** (.1578 AND .1908 compared to .306 and .369 from mixed model) this is related to incorrect assumption of the iid model that there are 7115 independent measures for these variables rather than only 160 (schools). On the other hand the standard error from the iid model for the individual level covariates, e.g. cses is not far off from the Mixed or GEE model (.1552 compared to .1551 from mixed model). This is due to there being relatively little school to school variability not accounted for as compared to the individual variability.