Models In Epidemiology And Biostatistics
Gordon Hilton Fick

Conditional Models With Repeated Measures

In considering the dental data, the asthma data and the pain data, we have identified a number of issues that need to be addressed.

1) Repeated measurements from the same individual are not independent. For example:

```
. use pott.dta
. reshape wide
. corr dist1 dist2 dist3 dist4
(obs=27)

             |    dist1    dist2    dist3    dist4
-------------+------------------------------------
       dist1 |   1.0000
       dist2 |   0.6256   1.0000
       dist3 |   0.7108   0.6349   1.0000
       dist4 |   0.5998   0.7593   0.7950   1.0000
```

2) Regression analyses adjusting for subjects has an analog with viewing subjects like confounders but subjects are not 'populations'. Rather they might be viewed as a sample from a population. We tried:

$E(y)=\beta_0+\sum_{j=2}^{27}\beta_j\delta_j+\beta_{28}\text{age}+\beta_{29}\text{age*sex}$ where δ_j is the indicator for the jth subject. The first 27 β_j attempt to 'adjust' for subjects. A more realistic plan conceives of u_j instead of β_j where the u_j are a sample from a population of such children.

3) We need to consider data that have varying numbers of repeated measurements per individual.
4) We need to consider data with varying values for the covariates. For example, in the dental data, we should use the child's actual age when each distance measurement was recorded.
5) The regression analyses considered so far have not enabled the separation of key between subjects comparisons. In the asthma data, we could not consider order modification in our regression approach.
6) So far, we have seen that the data could be placed in either wide format or in long format. In long format, a covariate of interest [age in the dental data] is recorded as a column with multiple rows per individual. In wide format, the covariate is implicitly a part of the outcome variable name. This would be acceptable with the data recorded in 'waves' [as in ages 8, 10 12, 14] but would not be practical if we recorded the child's actual age.
7) We would like to have methods that extend these matters to dichotomous outcomes [logistic regression and others], ordinal outcomes [proportional odds models and others], count outcomes [poisson, negative binomial and variants] and interval 'measured' outcomes [linear regression and its many extensions]
8) We need to identify other settings in which a lack of independence is crucial [matching, blocking, clustering; for example] and then anticipate that our methods extend to these settings.

We can write our models in vector/matrix notation:

$\boldsymbol{y}=X\boldsymbol{\beta}+\boldsymbol{\epsilon}$ [linear]

$\log(\frac{\boldsymbol{p}}{\mathbf{1}-\boldsymbol{p}}) = X\boldsymbol{\beta}$ [logistic: $\boldsymbol{p} = Pr(\boldsymbol{y}=\mathbf{1})$]

$\log\boldsymbol{\lambda} = X\boldsymbol{\beta} + \log \boldsymbol{e}$ [poisson or negative binomial with exposure $\boldsymbol{e}$; $E(\boldsymbol{y}) = \boldsymbol{\lambda}$

$\log(\frac{\boldsymbol{p}_j}{\mathbf{1}-\boldsymbol{p}_j}) = X\boldsymbol{\beta} - \boldsymbol{\kappa}_j$ [proportional odds model] $\boldsymbol{p}_j = Pr(\boldsymbol{y} > cut\ j)$

All our models use the assumption that $\boldsymbol{y}$ are statistically independent.

If $\boldsymbol{y}$ is normally distributed, statistical independence is the same as $VAR\ \boldsymbol{y}$ being a diagonal matrix. In many circumstances, we restrict this assumption even more and say that $VAR\ \boldsymbol{y} = \sigma^2 I$

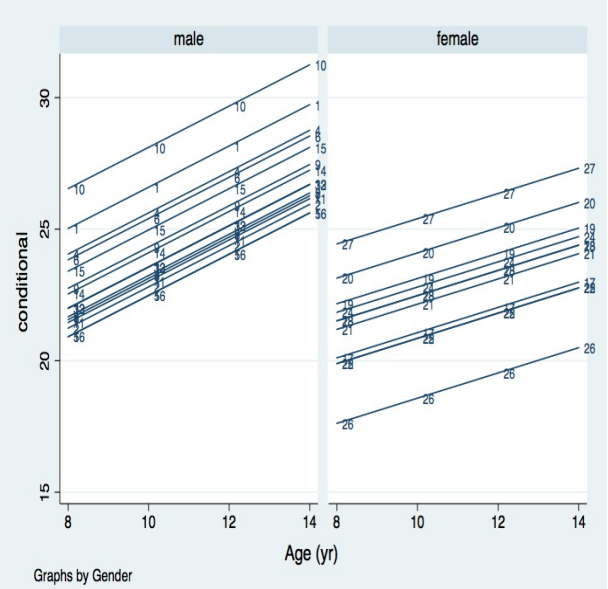
Let us now consider the 'simplest' conditional model:

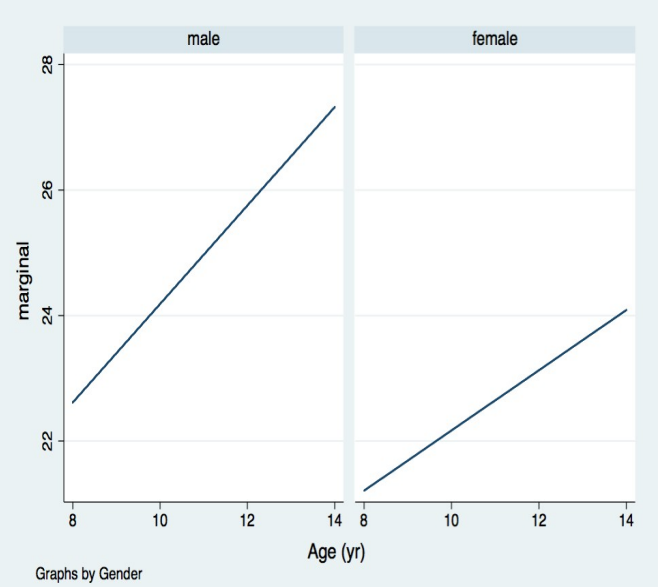
$\boldsymbol{y} = X\boldsymbol{\beta} + Z\boldsymbol{u} + \boldsymbol{\epsilon}$ or $E(\boldsymbol{y} : \boldsymbol{u}) = X\boldsymbol{\beta} + Z\boldsymbol{u}$ and, for now, we will assume that $VAR\ \boldsymbol{\epsilon} = \sigma^2 I$

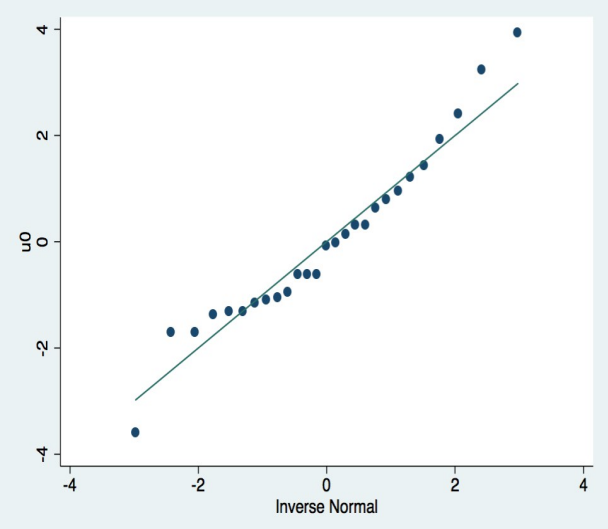
and let us return to the dental data:
$\boldsymbol{y}$ will be a 108 x 1 column vector; X will be a 108x4 matrix; $\boldsymbol{\beta}$ will be a 4x1 column vector;
$\boldsymbol{u}$ will be a 27x1 column vector; Z will be 108x27 matrix and $\boldsymbol{\epsilon}$ will be a 108x1 column vector.
The columns of Z are the indicators for each of the 27 subjects. The $\boldsymbol{u}$ vector is something new. They are not regression coefficients in the sense we seen before. They are unknown values from a probability distribution with mean zero and $VAR\ \boldsymbol{u} = \sigma_u^2 I$. Also, it is standard here to require that $\mathbf{1}'\boldsymbol{u} = 0$. This constraint ensures the correct interpretation of $\boldsymbol{\beta}$
With this model and our data, we will have estimates for the regression coefficients and σ^2 as before . Now we will get an estimate for σ_u^2 and predictions for the $\boldsymbol{u}$.
We then get different sets of quantities to study. We get estimates of $X\boldsymbol{\beta}$ like before but we now also get estimates of $E(\boldsymbol{y} : \boldsymbol{u}) = X\boldsymbol{\beta} + Z\ \boldsymbol{u}$

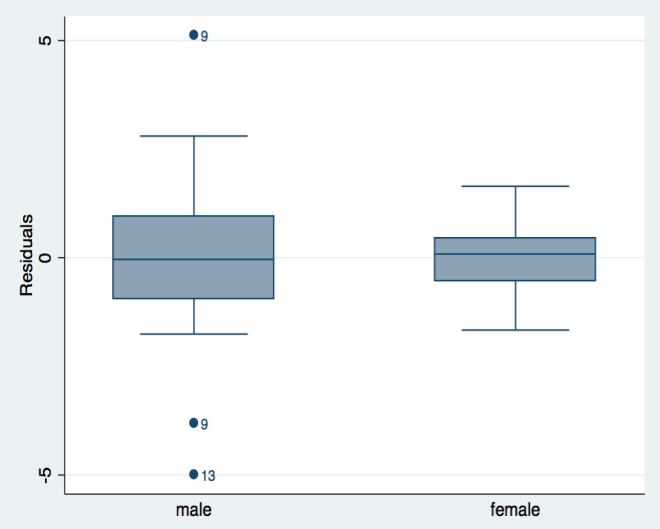
```
. mixed dist age sex as || subject:

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood = -214.31953
Iteration 1:   log likelihood = -214.31953

Computing standard errors:

Mixed-effects ML regression                     Number of obs      =       108
Group variable: subject                         Number of groups   =        27

                                                Obs per group: min =         4
                                                               avg =       4.0
                                                               max =         4


                                                Wald chi2(3)       =    142.05
Log likelihood = -214.31953                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        dist |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |    .784375   .0765383    10.25   0.000     .6343626    .9343874
```

```
         sex |   1.032102   1.508864     0.68   0.494    -1.925217    3.989422
          as |  -.3048295   .1199125    -2.54   0.011    -.5398537   -.0698054
       _cons |   16.34062   .9630849    16.97   0.000     14.45301    18.22824
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.      [95% Conf. Interval]
-----------------------------+------------------------------------------------
subject: Identity            |
                  var(_cons) |   3.030561   .9552074      1.633923    5.621014
-----------------------------+------------------------------------------------
               var(Residual) |   1.874597   .2945645      1.377699    2.550711
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =    49.60 Prob >= chibar2 = 0.0000

. predict yhm,xb
. predict yhc,fitted
. predict u, reffects
. predict res,res
. scatter yhc age,connect(ascending) msymbol(i) mlabel(subject) by(sex)
ytitle("conditional")

. scatter yhm age,connect(ascending) msymbol(i) by(sex) ytitle("marginal")
. gen u0=u
. by subject: replace u0=. if _n!=1
. qnorm u,ytitle("sorted observed")
. graph box res, over(sex) marker(1,mlabel(subject))
```









There are many ways to extend the above model. Lets try:
$E(\boldsymbol{y}:\boldsymbol{u_0}\boldsymbol{u_1})=X\beta+Z_0\boldsymbol{u_0}+Z_1\boldsymbol{u_1}$ where Z_0 contains the indicators for the subjects as before and Z_1 is again a 108x27 matrix that contains the ages for each child. Each row of [$\boldsymbol{u_0}\ \boldsymbol{u_1}$] then provides intercept and slope adjustment of each child. Further it fairly standard to model $VAR\ \boldsymbol{u_0}=\sigma^2_{u0}I$ and $VAR\ \boldsymbol{u_1}=\sigma^2_{u1}I$ and to allow for a covariance σ_{u0u1} . It is known that intercept and slope are typically correlated.

```
. mixed dist age sex as || subject: age,covariance(unstr)

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood = -213.92903
Iteration 1:   log likelihood = -213.90624
Iteration 2:   log likelihood = -213.90298
Iteration 3:   log likelihood = -213.90298

Computing standard errors:

Mixed-effects ML regression                     Number of obs      =       108
Group variable: subject                         Number of groups   =        27

                                                Obs per group: min =         4
                                                               avg =       4.0
                                                               max =         4


                                                Wald chi2(3)       =    121.59
Log likelihood = -213.90298                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        dist |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |    .784375   .0827532     9.48   0.000     .6221818    .9465682
         sex |   1.032102   1.535496     0.67   0.501    -1.977414    4.041618
          as |  -.3048295   .1296493    -2.35   0.019    -.5589375   -.0507216
       _cons |   16.34062   .9800834    16.67   0.000      14.4197    18.26155
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
subject: Unstructured        |
                   var(age) |   .0237593   .0340875      .0014276    .3954176
                 var(_cons) |   4.556941   4.671805       .610965    33.98839
              cov(age,_cons) |  -.1982569   .3790498     -.9411809    .5446671
-----------------------------+------------------------------------------------
               var(Residual) |   1.716202   .3302811      1.176944    2.502541
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(3) =    50.44   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.
```

The four graphs in the first example should be repeated.

```
. drop yhm yhc u u0 res
. predict yhm,xb
. predict yhc,fitted
. predict u1 u0, reffects
. predict res,res
. scatter yhc age,connect(ascending) msymbol(i) mlabel(subject) by(sex)
ytitle("conditional")
```

```
. scatter yhm age,connect(ascending) msymbol(i) by(sex) ytitle("marginal")
. by subject: replace u0=. if _n!=1
. by subject: replace u1=. if _n!=1
. qnorm u0,ytitle("sorted observed")
. qnorm u1,ytitle("sorted observed")
. scatter u1 u0
. graph box res, over(sex) marker(1,mlabel(subject))
```

Next, lets briefly look at the asthma data:

```
use forsal.dta
reshape long
gen ord =tr
replace ord =1-ord if grp==1
mixed pef tr ord grp ||id:

Mixed-effects ML regression                     Number of obs      =        24
Group variable: id                              Number of groups   =        12

                                                Obs per group: min =         2
                                                               avg =       2.0
                                                               max =         2


                                                Wald chi2(3)       =     19.11
Log likelihood = -121.55306                     Prob > chi2        =    0.0003

------------------------------------------------------------------------------
         pef |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          tr |  -38.33333   9.084224    -4.22   0.000    -56.13808   -20.52858
         ord |   3.333333   9.084224     0.37   0.714    -14.47142    21.13808
         grp |  -29.16667   26.93013    -1.08   0.279    -81.94875    23.61541
       _cons |   364.1667    20.0967    18.12   0.000     324.7779    403.5555
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Identity                 |
                  var(_cons) |   1928.126   893.9558      777.1136    4783.946
-----------------------------+------------------------------------------------
               var(Residual) |   495.1387   202.1395      222.4464    1102.119
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =    12.03 Prob >= chibar2 = 0.0003
```

Notice the resemblance [closeness] to the three t-tests considered earlier. Not exactly the same. The three t-tests perhaps get the edge [for this little study] since they correctly use the t-distribution for the p-values and confidence intervals. Our new conditional model uses approximate z values. On the other hand our conditional model approach is quite general and extensions to a wide range of cross over type studies are available.

Now lets revisit the pain study with conditional models:

$\text{logit}(\boldsymbol{p}:\boldsymbol{u}) = X\boldsymbol{\beta} + Z\,\boldsymbol{u}$ where $\boldsymbol{p} = \Pr(\boldsymbol{y} = \mathbf{1}:\boldsymbol{u})$

The regression coefficients now have interpretations in terms of conditional log odds [conditional on $\boldsymbol{u}$]. We have adjusted for subject specific components.

We can construct estimates assuming $\boldsymbol{u} = \mathbf{0}$ but these are not marginal log odds. We can construct predictions of the conditional log odds and the $\boldsymbol{u}$. Assessment of the $\boldsymbol{u}$ can be limited if there are only a fixed small set of possible outcomes and explanatory variables.

```
use pain.dta
replace grp=grp-1
gen trg=tr*grp
melogit imp tr grp trg || id:

Mixed-effects logistic regression               Number of obs      =       168
Group variable:              id                 Number of groups   =        84

                                                Obs per group: min =         2
                                                               avg =       2.0
                                                               max =         2

Integration method: mvaghermite                 Integration points =         7

                                                Wald chi2(3)       =     14.57
Log likelihood = -105.41698                     Prob > chi2        =    0.0022
------------------------------------------------------------------------------
         imp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          tr |   .3327301   .4738805     0.70   0.483    -.5960587    1.261519
         grp |  -1.530021   .5735093    -2.67   0.008    -2.654079   -.4059636
         trg |   1.891031    .752978     2.51   0.012     .4152214    3.366841
       _cons |   .2166222   .3527853     0.61   0.539    -.4748242    .9080687
-------------+----------------------------------------------------------------
id           |
   var(_cons)|   .6153444   .7999041                      .0481522    7.863574
------------------------------------------------------------------------------
LR test vs. logistic regression: chibar2(01) =     0.91 Prob>=chibar2 = 0.1704
```

For proportional odds, we now have:

$\text{logit}(\boldsymbol{p}_j : \boldsymbol{u}) = X\boldsymbol{\beta} - \kappa_j + Z\,\boldsymbol{u}$ where $\boldsymbol{p}_j = \Pr(\boldsymbol{y} > \text{cut}_j : \boldsymbol{u})$

Again, the regression coefficients are in terms of conditional log odds

```
. meologit ove tr grp trg || id:

Mixed-effects ologit regression                 Number of obs      =       168
Group variable:              id                 Number of groups   =        84

                                                Obs per group: min =         2
                                                               avg =       2.0
                                                               max =         2

Integration method: mvaghermite                 Integration points =         7

                                                Wald chi2(3)       =     24.12
Log likelihood = -251.87999                     Prob > chi2        =    0.0000
------------------------------------------------------------------------------
         ove |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          tr |   .0601627   .4147691     0.15   0.885    -.7527698    .8730952
         grp |  -1.528548    .525087    -2.91   0.004    -2.557699   -.4993961
         trg |   2.206275   .6230285     3.54   0.000     .9851615    3.427388
-------------+----------------------------------------------------------------
       /cut1 |  -5.220747   .7941854    -6.57   0.000    -6.777322   -3.664173
       /cut2 |  -1.396146   .4046063    -3.45   0.001     -2.18916   -.6031321
       /cut3 |  -.3287973    .380997    -0.86   0.388    -1.075538    .4179431
       /cut4 |   .5194378   .3821936     1.36   0.174    -.2296478    1.268523
       /cut5 |   2.978362    .511053     5.83   0.000     1.976716    3.980007
-------------+----------------------------------------------------------------
id           |
   var(_cons)|   1.498867   .7778412                      .5420376    4.144733
------------------------------------------------------------------------------
LR test vs. ologit regression:   chibar2(01) =     7.33 Prob>=chibar2 = 0.0034
```

melogit provides the conditional model for repeated dichotomous outcomes and the logit link.

meologit provides the conditional model for repeated ordinal outcomes, the logit link and the proportional odds assumption.

Now let us consider a study of count outcomes.

$\log(\boldsymbol{\lambda}:\boldsymbol{u}) = X\boldsymbol{\beta} + Z\,\boldsymbol{u}$ where $\boldsymbol{y} \sim \text{poisson}(\boldsymbol{\lambda}:\boldsymbol{u})$

The details are given in the paper by Breslow and Clayton (1993)

```
use epilepsy.dta
gen ls1=log(seizures+1)
sort subject vs
lab def trl 0 placebo 1 active
lab val treat trl
scatter ls1 vs,connect(ascending) by(treat) mlabel(subject)
table treat,c(mean lbas sd lbas)
table treat vs,c(mean ls1 sd ls1)
gen trv=treat*vs
mepoisson seizures treat vs trv || subject:

Mixed-effects Poisson regression                Number of obs      =       236
Group variable:         subject                 Number of groups   =        59

                                                Obs per group: min =         4
                                                               avg =       4.0
                                                               max =         4

Integration method: mvaghermite                 Integration points =         7

                                                Wald chi2(3)       =     10.26
Log likelihood = -695.90821                     Prob > chi2        =    0.0164
------------------------------------------------------------------------------
    seizures |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       treat |  -.2126636   .2713048    -0.78   0.433    -.7444113    .3190841
          vs |  -.0428052    .028845    -1.48   0.138    -.0993403    .0137299
         trv |  -.0314748   .0405755    -0.78   0.438    -.1110014    .0480518
       _cons |   1.878152   .1953233     9.62   0.000     1.495326    2.260979
-------------+----------------------------------------------------------------
subject      |
   var(_cons)|   .8788584   .1788546                      .5897836    1.309619
------------------------------------------------------------------------------
LR test vs. Poisson regression:  chibar2(01) =  1884.04 Prob>=chibar2 = 0.0000
. mepoisson seizures treat vs trv || subject: vs ,cov(unstr)

Mixed-effects Poisson regression                Number of obs      =       236
Group variable:         subject                 Number of groups   =        59

                                                Obs per group: min =         4
                                                               avg =       4.0
                                                               max =         4

Integration method: mvaghermite                 Integration points =         7

                                                Wald chi2(3)       =      3.30
Log likelihood = -686.33683                     Prob > chi2        =    0.3470
------------------------------------------------------------------------------
    seizures |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       treat |  -.2509616   .2987855    -0.84   0.401    -.8365704    .3346471
          vs |  -.0434861    .045908    -0.95   0.344     -.133464    .0464918
         trv |  -.0138791   .0627282    -0.22   0.825    -.1368241    .1090658
       _cons |   1.864557   .2166208     8.61   0.000     1.439987    2.289126
-------------+----------------------------------------------------------------
subject      |
      var(vs)|   .0211017   .0091714                      .0090024    .0494622
```

```
   var(_cons)|   1.032918    .243383                      .650879    1.639199
-------------+------------------------------------------------------------------
subject      |
cov(_cons,vs)|  -.0573591   .0381642    -1.50   0.133    -.1321596    .0174414
------------------------------------------------------------------------------
LR test vs. Poisson regression:      chi2(3) =  1903.18   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.
```

$\log(\boldsymbol{\lambda}:\boldsymbol{u}) = X\boldsymbol{\beta} + Z\,\boldsymbol{u}$ where $\boldsymbol{y}:\boldsymbol{u} \sim \mathrm{NB}(\boldsymbol{\lambda})$ with $E(\boldsymbol{y}:\boldsymbol{u}) = \boldsymbol{\lambda}$ and $VAR(\boldsymbol{y}:\boldsymbol{u}) = diag[\boldsymbol{\lambda}(1+\alpha\boldsymbol{\lambda})]$

```
. menbreg seizures treat vs trv || subject:

Mixed-effects nbinomial regression              Number of obs      =       236
Overdispersion:            mean
Group variable:         subject                 Number of groups   =        59

                                                Obs per group: min =         4
                                                               avg =       4.0
                                                               max =         4

Integration method: mvaghermite                 Integration points =         7

                                                Wald chi2(3)       =      3.97
Log likelihood = -655.11664                     Prob > chi2        =    0.2649
------------------------------------------------------------------------------
    seizures |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       treat |  -.2438501   .3021013    -0.81   0.420    -.8359577    .3482575
          vs |  -.0447497   .0465564    -0.96   0.336    -.1359985    .0464991
         trv |  -.0184578   .0667353    -0.28   0.782    -.1492566     .112341
       _cons |    1.89958   .2157934     8.80   0.000     1.476633    2.322528
-------------+----------------------------------------------------------------
    /lnalpha |  -2.008753   .2353817    -8.53   0.000    -2.470092   -1.547413
-------------+----------------------------------------------------------------
subject      |
   var(_cons)|   .8447953   .1788146                      .5579315    1.279152
------------------------------------------------------------------------------
LR test vs. nbinomial regression:chibar2(01) =   183.61 Prob>=chibar2 = 0.0000
```

Mean dispersion is the default. Constant dispersion is an option.

Breslow & Clayton provide one analysis and discussion. The following is a close version of their results.

```
. mepoisson seizures lbas treat lbas_trt lage v4 vs || subject: vs ,cov(unstr)

Mixed-effects Poisson regression                Number of obs      =       236
Group variable:         subject                 Number of groups   =        59

                                                Obs per group: min =         4
                                                               avg =       4.0
                                                               max =         4

Integration method: mvaghermite                 Integration points =         7

                                                Wald chi2(6)       =    118.18
Log likelihood = -654.32718                     Prob > chi2        =    0.0000
------------------------------------------------------------------------------
    seizures |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lbas |   .8850952   .1312167     6.75   0.000     .6279152    1.142275
       treat |  -.9281649   .4020194    -2.31   0.021    -1.716108   -.1402214
    lbas_trt |   .3378666   .2043677     1.65   0.098    -.0626867    .7384199
       lage |   .4760858    .353526     1.35   0.178    -.2168124    1.168984
          v4 |  -.1392905   .0845376    -1.65   0.099    -.3049812    .0264001
```

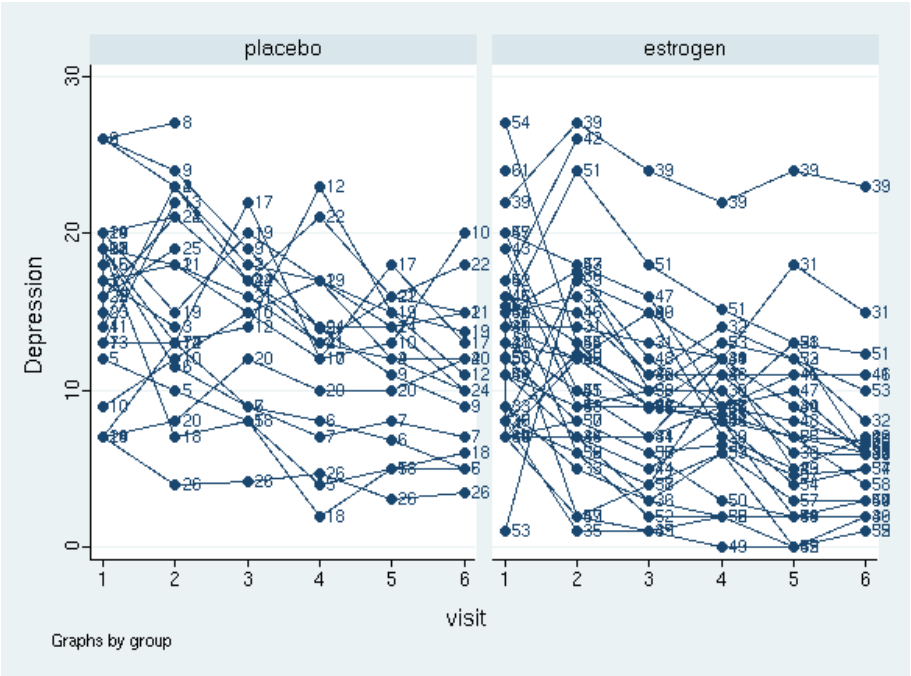
```
          vs |  -.0131428   .0415294    -0.32   0.752    -.0945388    .0682532
       _cons |   2.165474   .2410654     8.98   0.000     1.692995    2.637954
-------------+------------------------------------------------------------------
subject      |
      var(vs)|   .0230964   .0098286                      .0100304    .0531828
   var(_cons)|   .3929233   .1249306                      .2107011    .7327381
-------------+------------------------------------------------------------------
subject      |
cov(_cons,vs)|  -.0571615   .0308857    -1.85   0.064    -.1176965    .0033734
------------------------------------------------------------------------------
LR test vs. Poisson regression:      chi2(3) =   325.78   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.
```

Lets now take a look at a study in Gregoire (1996). We can look at the data with:

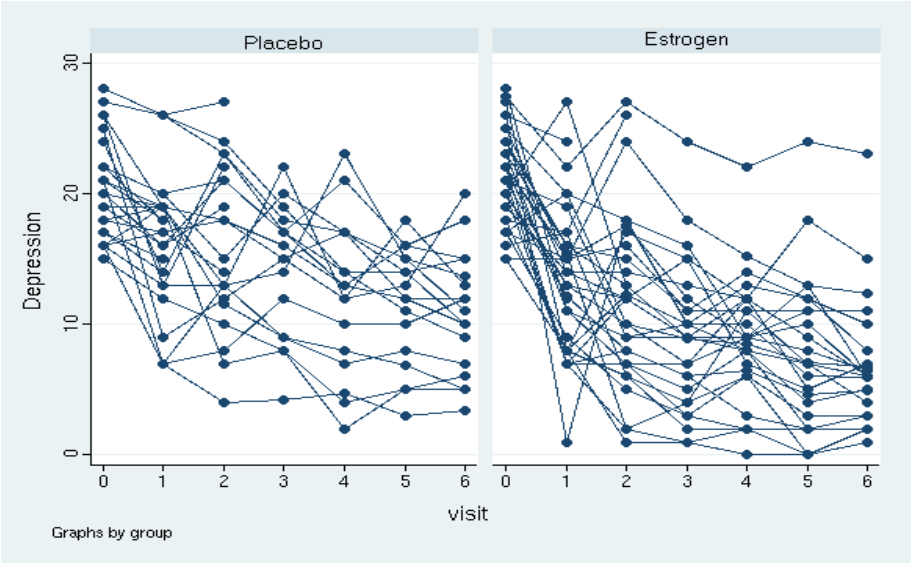
```
use depress.dta
sort group subj visit
twoway connected dep visit,connect(ascending) by(group) ytitle(Depression) xlabel(1/6)
```

This data set display is probably too busy and we should, perhaps, regraph with about 10 subjects or so per display. Nevertheless, this display illustrates a whole range of issues that we did not see in the dental study. Some subjects have only one score after baseline (like subject #61) . Some subjects have only 2 scores (like subject #8). The majority of subjects completed all 6 visits. Subject #39 seems so different from the rest. Subject #53 appears to have an "impossibly" low score at visit #1 especially as this person's pre-assignment score was 26. Should we treat the 'completers' the same as the 'dropouts'? Also, there are ten dropouts in the placebo and six in the active group.

```
. table group visit,c(n dep)

-----------------------------------------------
          |                visit
    group |    1     2     3     4     5     6
----------+------------------------------------
  placebo |   27    22    17    17    17    17
 estrogen |   34    31    29    28    28    28
-----------------------------------------------
```

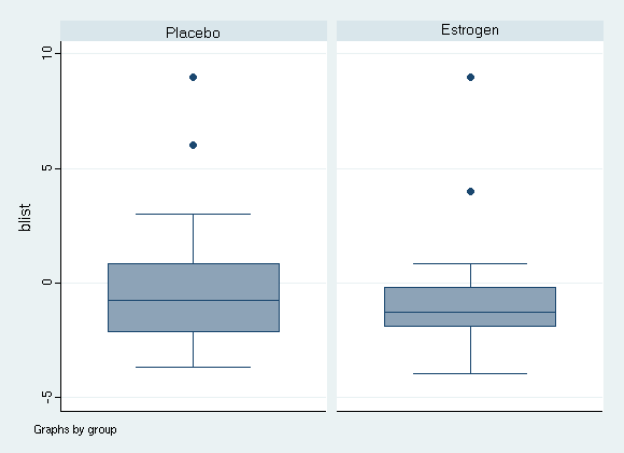


Here is the same data with the baseline values added as 'visit 0':



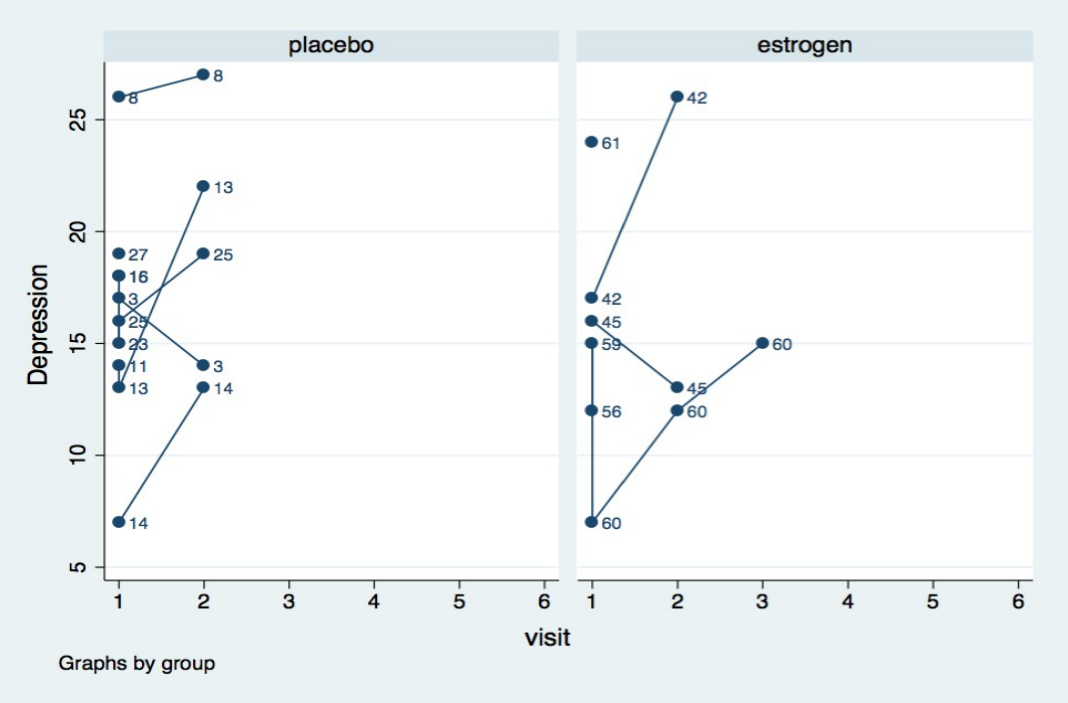
Notice that those receiving placebo seem to 'benefit' from being in this study. Placebos can be wonderful things :-) Lets further explore the idea of a post baseline slope as we did with the dental study. This is a bit trickier still in that we need to advise Stata to ignore those individuals with only one post baseline measurement. This requires use of the 'capture' command and the use of a system variable _rc. Have a look at the Stata Programming Manual for more on this valuable and powerful set of utilities.

```
quietly forval num = 1/61 {
capture regr dep visit if subj==`num'
if _rc==0 regr dep visit if subj==`num'
if _rc==0 replace blist=_b[visit] if subj==`num'
}
by subj: replace blist=. if _n!=1
```



This display shows that median slope may be more negative for those receiving estrogen but there are serious outliers [that come from individuals with only 2 measurements]. Here is a display of the dropouts

```
twoway connected dep visit if drop==1,connect(ascending) by(group) ytitle(Depression)
xlabel(1/6) mlabel(subj)
```



The next two analyses enable us to consider dropout status as potential modifier/confounder. Completers: drop==0; dropouts: drop==1.

```
. mixed dep group drop dg || subj:
Mixed-effects ML regression                     Number of obs      =       295
Group variable: subj                            Number of groups   =        61

                                                Obs per group: min =         1
                                                               avg =       4.8
                                                               max =         6


                                                Wald chi2(3)       =     26.94
Log likelihood = -876.67365                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         dep |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |  -4.075945   1.308373    -3.12   0.002     -6.64031   -1.511581
        drop |   3.766207   1.934725     1.95   0.052    -.0257844    7.558199
          dg |   3.115024   2.957592     1.05   0.292    -2.681749    8.911798
       _cons |   13.37725   1.032059    12.96   0.000     11.35446    15.40005
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.      [95% Conf. Interval]
-----------------------------+------------------------------------------------
subj: Identity               |
                  var(_cons) |   15.46552   3.469955       9.96283    24.00746
-----------------------------+------------------------------------------------
               var(Residual) |    15.8517   1.453268      13.24459      18.972
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =   104.98 Prob >= chibar2 = 0.0000
```

```
. mixed dep group drop || subj:
Mixed-effects ML regression                     Number of obs      =       295
Group variable: subj                            Number of groups   =        61

                                                Obs per group: min =         1
                                                               avg =       4.8
                                                               max =         6


                                                Wald chi2(2)       =     25.54
Log likelihood = -877.22587                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         dep |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |  -3.464283   1.180574    -2.93   0.003    -5.778165     -1.1504
        drop |   5.100048   1.470863     3.47   0.001     2.217209    7.982886
       _cons |   12.99666   .9732443    13.35   0.000     11.08914    14.90419
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
subj: Identity               |
                  var(_cons) |   15.69568   3.529713      10.10086    24.38945
-----------------------------+------------------------------------------------
               var(Residual) |   15.87724    1.45723      13.26328    19.00637
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =   105.01 Prob >= chibar2 = 0.0000
```

Now we can consider the distribution of the residuals ϵ conditional on the subject specific components u . The default is to assume [conditional] independence and constant variance. There are a very wide range of possible forms for this distribution. We do continue to assume independence between the residuals from different subjects.
As a couple of examples, we can allow for possibly different variances, say, between treatment groups or allow for the Toeplitz form of covariances.

```
. mixed dep group drop || subj:,res(,by(group))

Mixed-effects ML regression                     Number of obs      =       295
Group variable: subj                            Number of groups   =        61

                                                Obs per group: min =         1
                                                               avg =       4.8
                                                               max =         6


                                                Wald chi2(2)       =     25.57
Log likelihood =    -877.22                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         dep |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |  -3.464819   1.180137    -2.94   0.003    -5.777845   -1.151792
        drop |   5.094085   1.469908     3.47   0.001     2.213119    7.975051
       _cons |   12.99742   .9723903    13.37   0.000     11.09157    14.90327
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
subj: Identity               |
                  var(_cons) |    15.6917   3.528878      10.09822    24.38345
```

```
-----------------------------+------------------------------------------------
Residual: Independent,       |
    by group                 |
           placebo: var(e) |   15.68204   2.291416       11.7768    20.88227
          estrogen: var(e) |   16.00389   1.884266      12.70594    20.15785
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(2) =   105.02   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.
. mixed dep group drop || subj:,res(toeplitz 1,t(visit))

Mixed-effects ML regression                     Number of obs      =       295
Group variable: subj                            Number of groups   =        61

                                                Obs per group: min =         1
                                                               avg =       4.8
                                                               max =         6


                                                Wald chi2(2)       =     25.94
Log likelihood = -856.75753                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         dep |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |  -3.504549   1.145759    -3.06   0.002    -5.750195   -1.258903
        drop |   5.004422   1.460866     3.43   0.001     2.141178    7.867666
       _cons |   13.08391   .9416329    13.89   0.000     11.23834    14.92948
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.      [95% Conf. Interval]
-----------------------------+------------------------------------------------
subj: Identity               |
                  var(_cons) |   12.20801    3.39303      7.080542    21.04861
-----------------------------+------------------------------------------------
Residual: Toeplitz(1)        |
                       cov1 |   6.957426   1.199411      4.606623    9.308229
                     var(e) |   18.07575   1.922631      14.67432    22.26561
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(2) =   145.95   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

.  estat wcorr

Standard deviations and correlations for subj = 1:

Standard deviations:

       visit |      1       2       3       4       5       6
-------------+------------------------------------------------
          sd |  5.503   5.503   5.503   5.503   5.503   5.503

Correlations:

       visit |      1       2       3       4       5       6
-------------+------------------------------------------------
           1 |  1.000
           2 |  0.633   1.000
           3 |  0.403   0.633   1.000
           4 |  0.403   0.403   0.633   1.000
           5 |  0.403   0.403   0.403   0.633   1.000
           6 |  0.403   0.403   0.403   0.403   0.633   1.000
```

```
. estat wcorr,cov

Covariances for subj = 1:

       visit |      1       2       3       4       5       6
-------------+------------------------------------------------
           1 | 30.284
           2 | 19.165  30.284
           3 | 12.208  19.165  30.284
           4 | 12.208  12.208  19.165  30.284
           5 | 12.208  12.208  12.208  19.165  30.284
           6 | 12.208  12.208  12.208  12.208  19.165  30.284
```