<u>Assignment 3</u>
Plots of Repeated Measures Longitudinal Data

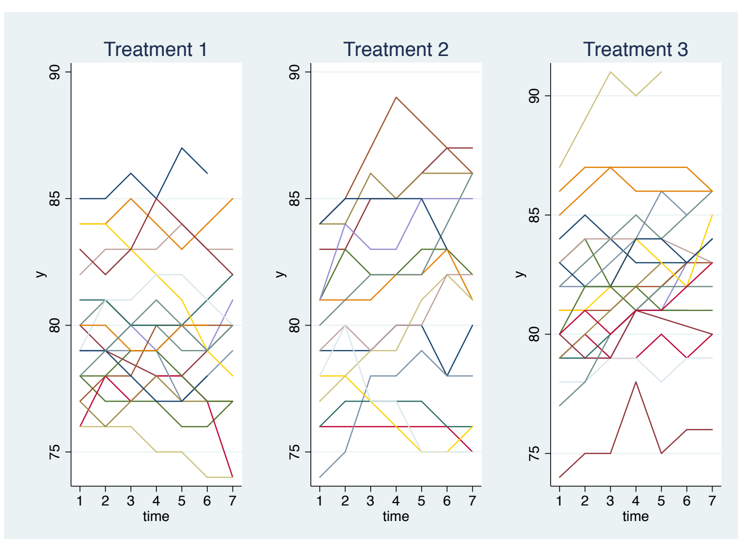
This homework is intended to develop skills in representing repeated measures data graphically (as well as a small lesson on the permutation test). Download the relevant data set and then complete the following tasks and answer the questions.

<u>Strength Data</u>
Trial data consisting of 3 treatment groups: No training (tx=1), weight training with light weights and high repetition (tx=2), and weight training with heavy weights and low repetition (tx=3). Subjects were followed for 7 weeks and a measure of muscle strength was recorded each week. The questions of interest are

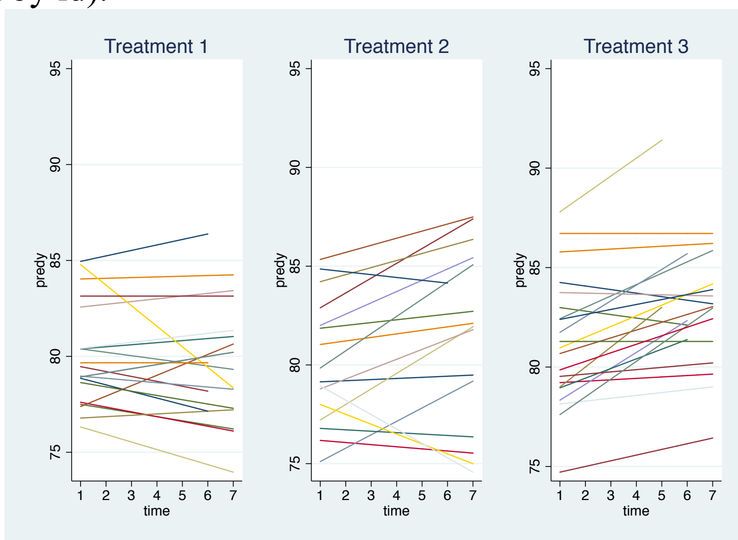
- Does weight training have any impact on strength?
- Is there a difference between tx 2 and 3?
- Which training program works quickest to increase strength?

The assignment is to address these using graphical methods and a statistical approach. Tip: A great deal of Stata syntax for graphics is covered in the Chapter 2 lecture slides and recent lab materials.

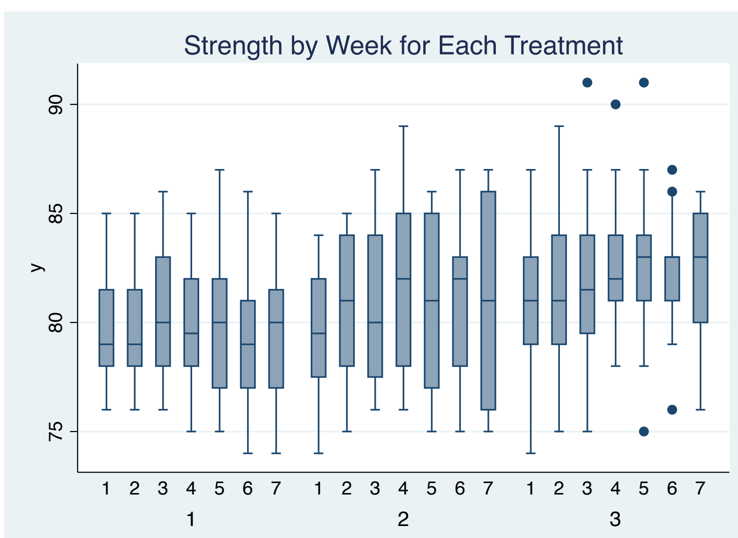
1. On one plot, plot separate line plots of all subjects' trajectories of strength versus week (relevant commands: *xtline, graph combine*).



2. Fit a linear regression by subject, and repeat 1, replacing the strength with predicted strength based on these regressions (relevant commands: *xtline, graph combine,* along with code below for running regression by id).



3. Optional: Consider whether there are other plots that might be useful, and include one more if you want to. (For example, box plots of the distribution of strength by both week and tx (relevant command *graph box* – note you can have more than one *over(var)* statement as option). Or feel free to choose another)



4. Write a short paragraph describing what the plots suggest about the bulleted questions above.

   From the first plot, it's hard to get any sense of whether the training regimes are affecting strength; the amount of overlap and ups and downs in the individual trajectories make it hard to see trends. In the plot of per-subject regression lines, it appears that in both treatments 2 and 3, most individuals are increasing in strength over the 7s weeks, though there's a great deal of variation in slops and some exceptions. Those in treatment 1 (the control) show no overall pattern. The boxplots obscure the individual trajectories but suggest that media strength measurements are increasing in both treatments 2 and 3. There is a greater spread of measurements for treatment 2, and the medias for treatment 3 are higher at both the beginning and end. It's hard to tell if treatment 2 or 3 works more quickly.

5. Reduce the data for each subject to 1 number – the slope of the change in strength estimated in each person separately (can use program below). One way you could test the treatment effect of tx=3 versus tx=2 on this outcome (ignoring tx=1) is to use a standard two-sample t-test.

```
. ttest slope1, by(tx1)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       2 |      16    .2313342    .1163581    .4654324   -.0166772    .4793457
       3 |      21    .3499837    .0807511    .3700481    .1815398    .5184276
---------+--------------------------------------------------------------------
combined |      37    .2986758    .0677611    .4121745      .16125    .4361017
---------+--------------------------------------------------------------------
    diff |           -.1186495    .1372595               -.3973012    .1600022
------------------------------------------------------------------------------
    diff = mean(2) - mean(3)                                      t =  -0.8644
Ho: diff = 0                                     degrees of freedom =       35

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.1966         Pr(|T| > |t|) = 0.3932          Pr(T > t) = 0.8034
```

   The ttest suggests there is no difference between treatments 2 and 3.

6. To further explore the relationship between treatment and group, try fitting a simple cross sectional model to this data as discussed in Chapter 3. Look the model with both naïve and robust standard errors. What happens if you expand the model to include a longitudinal term?

There are multiple ways to approach this question. The most straightforward is to add an interaction terms between treatment and time, after generating dummy variables for treatment. This allows you to test whether there's a difference between the slope on time for treatment groups 2 and 3. There is not, so we reach the same conclusion: we cannot distinguish whether treatment 2 or 3 is more effective, though both training regimes are better at increasing strength than the control (no training).

If you generate a longitudinal term for time, as we saw in lab, it will be collinear with the time variable.

```
. xtgee y _Itx_2 _Itx_3 time time_tx2 time_tx3, i(id) family(gaussian)

Iteration 1: tolerance = .12732624
Iteration 2: tolerance = .00003826
Iteration 3: tolerance = 6.453e-09

GEE population-averaged model                   Number of obs      =       370
Group variable:                         id      Number of groups   =        57
Link:                             identity      Obs per group: min =         5
Family:                           Gaussian                     avg =       6.5
Correlation:                  exchangeable                     max =         7
                                                Wald chi2(5)       =     64.83
Scale parameter:                  9.918406      Prob > chi2        =    0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      _Itx_2 |  -.1915965   1.039047    -0.18   0.854    -2.228091    1.844898
      _Itx_3 |    .956994   .9684046     0.99   0.323    -.9410442    2.855032
        time |  -.0693602   .0497936    -1.39   0.164    -.1669539    .0282334
    time_tx2 |   .3119225   .0739552     4.22   0.000      .166973     .456872
    time_tx3 |   .3730803   .0706584     5.28   0.000     .2345924    .5115681
       _cons |   80.06181   .6927472   115.57   0.000     78.70405    81.41957
------------------------------------------------------------------------------

. xtgee y _Itx_2 _Itx_3 time time_tx2 time_tx3, i(id) family(gaussian) robust

Iteration 1: tolerance = .12732624
Iteration 2: tolerance = .00003826
Iteration 3: tolerance = 6.453e-09

GEE population-averaged model                   Number of obs      =       370
Group variable:                         id      Number of groups   =        57
Link:                             identity      Obs per group: min =         5
Family:                           Gaussian                     avg =       6.5
Correlation:                  exchangeable                     max =         7
                                                Wald chi2(5)       =     33.82
Scale parameter:                  9.918406      Prob > chi2        =    0.0000

                                   (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |             Robust
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      _Itx_2 |  -.1915965   .9714351    -0.20   0.844    -2.095574    1.712381
      _Itx_3 |    .956994   .9449735     1.01   0.311      -.89512    2.809108
        time |  -.0693602   .0765871    -0.91   0.365    -.2194681    .0807477
    time_tx2 |   .3119225   .1357602     2.30   0.022     .0458374    .5780076
    time_tx3 |   .3730803   .1036701     3.60   0.000     .1698906      .57627
       _cons |   80.06181   .5997272   133.50   0.000     78.88636    81.23725
------------------------------------------------------------------------------
```

```
. lincom  time_tx2-time_tx3

 ( 1)  time_tx2 - time_tx3 = 0

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -.0611578   .1320877    -0.46   0.643    -.3200449    .1977294
------------------------------------------------------------------------------

.
end of do-file
```

Regression by id

```
** Initialize variables used in regbyid
capture drop predy
gen predy = .
capture drop slope
gen slope = .
capture drop intercept
gen intercept = .
** Program to fit regression by subject
capture program drop regbyid
program define regbyid, byable(recall)
syntax [varlist] [if] [in]
marksample touse
capture matrix drop beta
regress `varlist' if `touse'
matrix beta = get(_b)
replace slope = beta[1,1] if `touse'
replace intercept = beta[1,2] if `touse'
capture drop predt
predict predt
replace predy = predt if `touse'
end
** Sort by id
sort id
** Get Slope by subject
quietly by id: regbyid y time
```