1. **Objective:** analyzing epileptic seizures data using GEE marginal model in STATA.

2. **Scientific question**: Determine whether the treatment reduces the rate of epileptic seizures.

3. **Dataset:** epileptic seizures data set (on the class website)

   Data description: The number of epileptic seizures was recorded during a baseline period of eight weeks for 59 epileptics. Then the patients were randomized to treatment with the anti-epileptic drug progabide, or to placebo in addition to standard chemotherapy. The number of seizures was then recorded in four consecutive two-week intervals. The data set has the following five columns:

   Column 1: patient id number
   Column 2: seizures count in two-week interval
   Column 3: time
   Column 4: treatment group indicator (0=placebo, 1=treatment)
   Column 5: baseline seizures count in 8 weeks interval

   Outcome: seizures count, is a count data.
   Covariate: time and treatment.

4. STATA output of the analysis

- **read the data into STATA**

```
. log using c:\data\lab13, replace
. infile id count time treat baseline age using c:\data\seize.data.txt, clear
(236 observations read)
. save c:\data\seize, replace
file c:\data\seize.dta saved
```
.*short summary
```
. sum count baseline age

    Variable |      Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       count |      236    8.262712    12.35636         0        102
    baseline |      236    31.22034    26.70505         6        151
         age |      236    28.33898     6.26129        18         42
```

.*make transformations
```
. gen sqy=sqrt(count)
. *average to two weeks
. gen sqb=sqrt(baseline/4)
. *make the correction to zero count and take log transformation
. replace count=0.5 if count==0
(23 real changes made)
. gen ly=log(count)
. gen lb=log(baseline)
. *change to wide format
. reshape wide count ly sqy, i(id) j(time)
(note: j = 1 2 3 4)
Data                                   long   ->   wide
-----------------------------------------------------------------------
```

```
Number of obs.                          236    ->        59
Number of variables                      10    ->        18
j variable (4 values)                  time    ->     (dropped)
xij variables:
                                      count    ->     count1 count2 ... count4
                                         ly    ->     ly1 ly2 ... ly4
                                        sqy    ->     sqy1 sqy2 ... sqy4
```

- **<u>summarize the original counts</u>**

. sum baseline count1 count2 count3 count4

Variable	Obs	Mean	Std. Dev.	Min	Max
baseline	59	31.22034	26.87716	6	151
count1	59	8.983051	14.81494	.5	102
count2	59	8.389831	10.15996	.5	65
count3	59	8.508475	14.10841	.5	76
count4	59	7.364407	9.605059	.5	63

. sum baseline count1 count2 count3 count4 if treat==0

Variable	Obs	Mean	Std. Dev.	Min	Max
baseline	28	30.78571	26.10429	6	111
count1	28	9.375	10.12023	.5	40
count2	28	8.339286	8.109269	.5	29
count3	28	8.839286	14.64016	.5	76
count4	28	7.982143	7.609062	.5	29

. sum baseline count1 count2 count3 count4 if treat==1

Variable	Obs	Mean	Std. Dev.	Min	Max
baseline	31	31.6129	27.98175	7	151
count1	31	8.629032	18.21765	.5	102
count2	31	8.435484	11.84817	.5	65
count3	31	8.209677	13.84664	.5	72
count4	31	6.806452	11.20616	.5	63

. sort treat
. by treat: sum baseline age

-> treat = 0

Variable	Obs	Mean	Std. Dev.	Min	Max
baseline	28	30.78571	26.10429	6	111
age	28	29	6	19	42

-> treat = 1

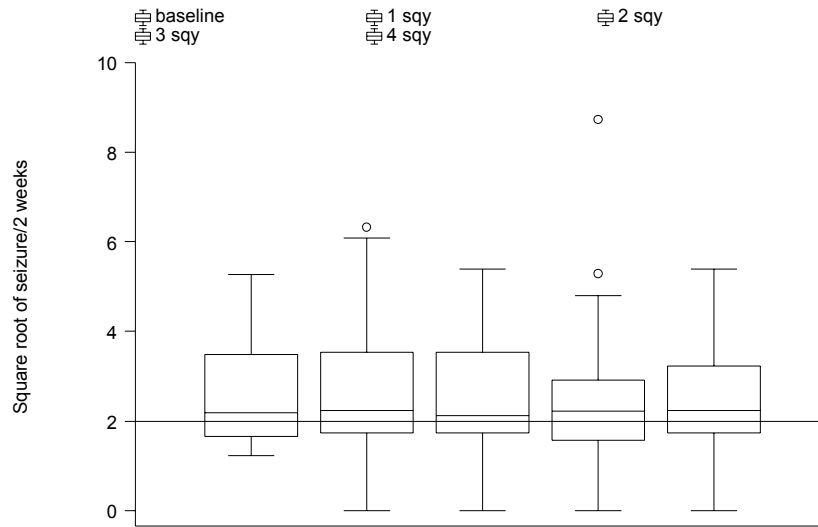| Variable	Obs	Mean	Std. Dev.	Min	Max
baseline	31	31.6129	27.98175	7	151
age	31	27.74194	6.602867	18	41

- **<u>some plots</u>**

. label var sqb "baseline"
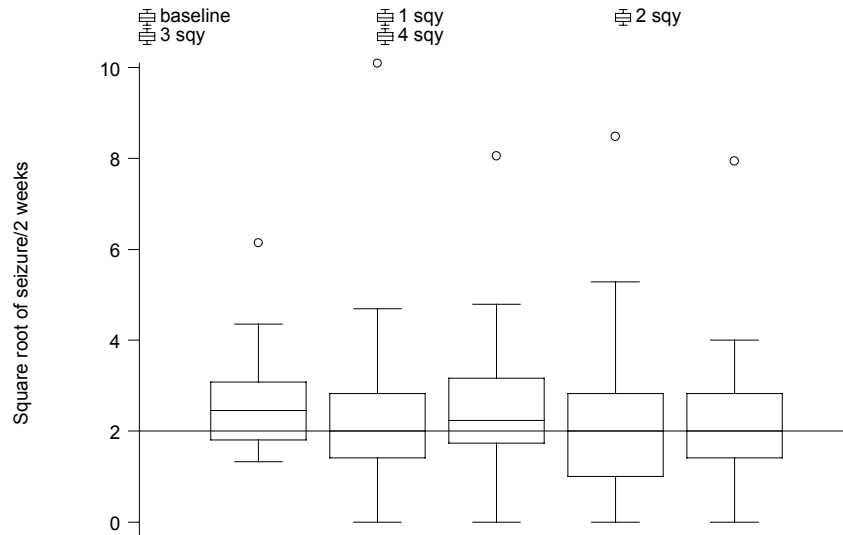. *box plot for placebo group
. graph sqb sqy1 sqy2 sqy3 sqy4 if treat==0, box ylab(0 2 to 10) yline(2)
l1title("Square root of seizure/2 weeks")

```
. *box plot for treatment group
. graph sqb sqy1 sqy2 sqy3 sqy4 if treat==1, box ylab(0 2 to 10) yline(2)
l1title("Square root of seizure/2 weeks")
```


The above two box plots suggest that a small reduction in the average number except possibly at week two.

```
. *change back to long format
. reshape long
(note: j = 1 2 3 4)

Data                              wide   ->   long
-----------------------------------------------------------------------------
Number of obs.                      59   ->     236
Number of variables                 18   ->      10
j variable (4 values)                    ->   time
xij variables:
            count1 count2 ... count4   ->   count
                  ly1 ly2 ... ly4      ->   ly
                sqy1 sqy2 ... sqy4     ->   sqy
-----------------------------------------------------------------------------
```
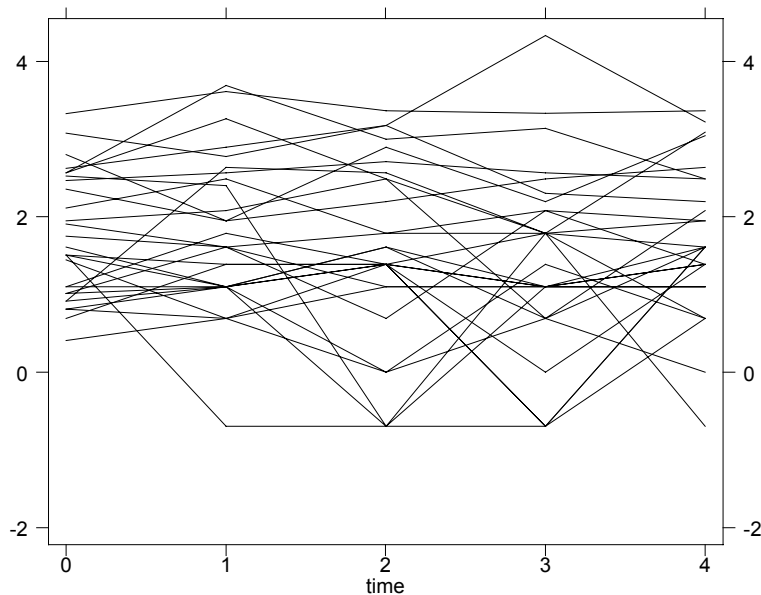
To include the baseline value in the graphs, we duplicate first observation for each subject (week=1) and change one of the duplicates for each subject to represent the baseline measure (week=0). Since the baseline measure represents seizure counts over an eight weeks interval, compared with two weeks interval for each of the other time points, we divide the baseline measure by 4.
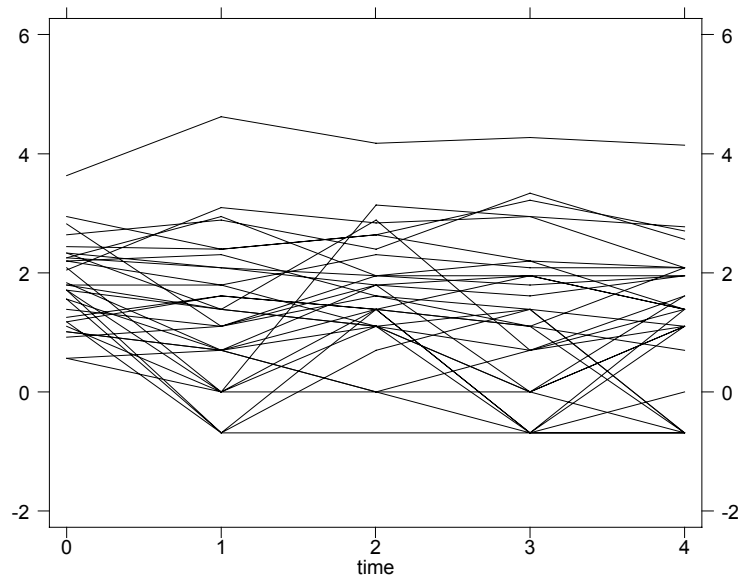
```
. expand 2 if time==1
(59 observations created)
. sort id time
. quiet by id: replace time=0 if _n==1
. replace ly =log(baseline/4) if time==0
(59 real changes made)
```

Since we plan to fit a Poisson model with the log link to the data, we look at the log transformation data when plotting the response profiles.

```
. *spaghetti plots
. graph ly time if treat==0, c(L) s(i) xlab ylab rlabel
```
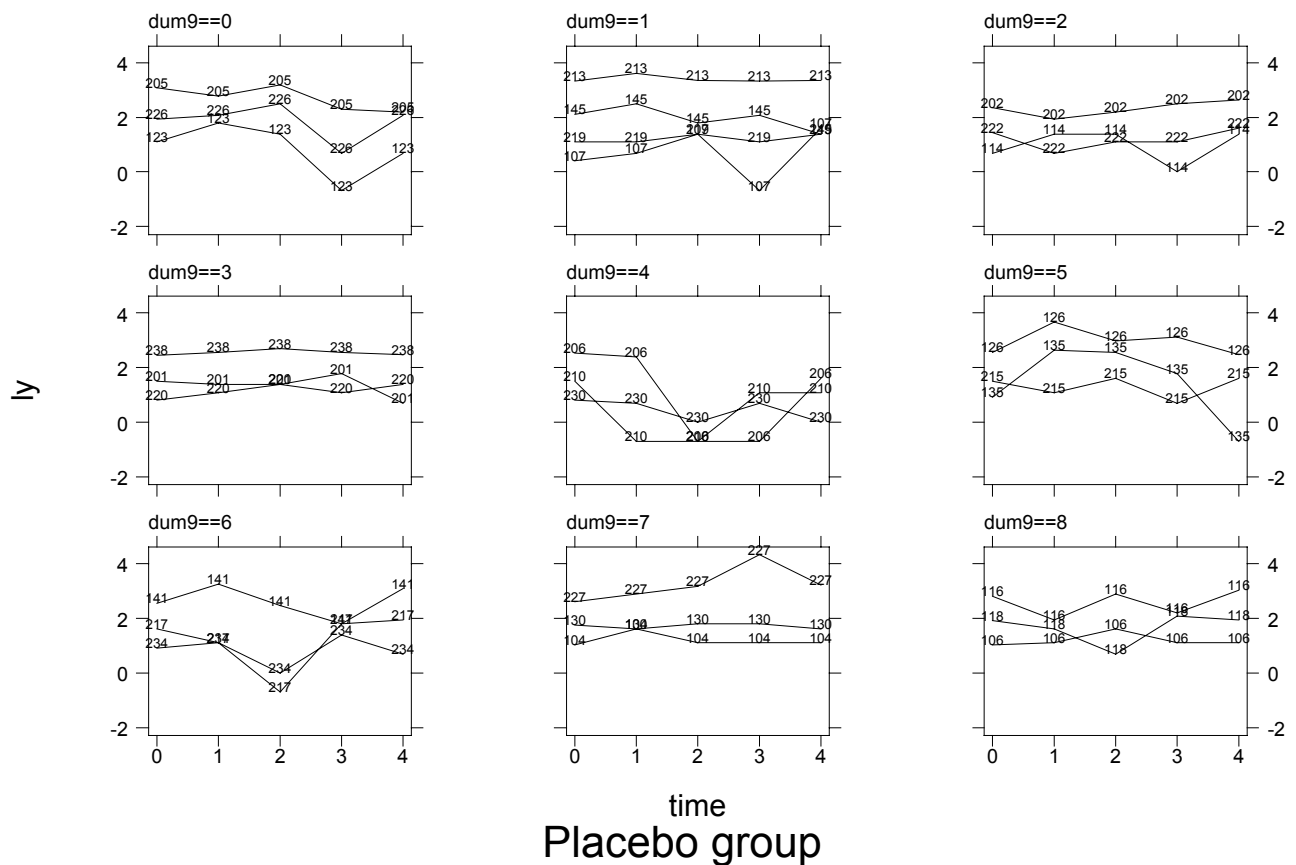


```
. graph ly time if treat==1, c(L) s(i) xlab ylab rlabel
```
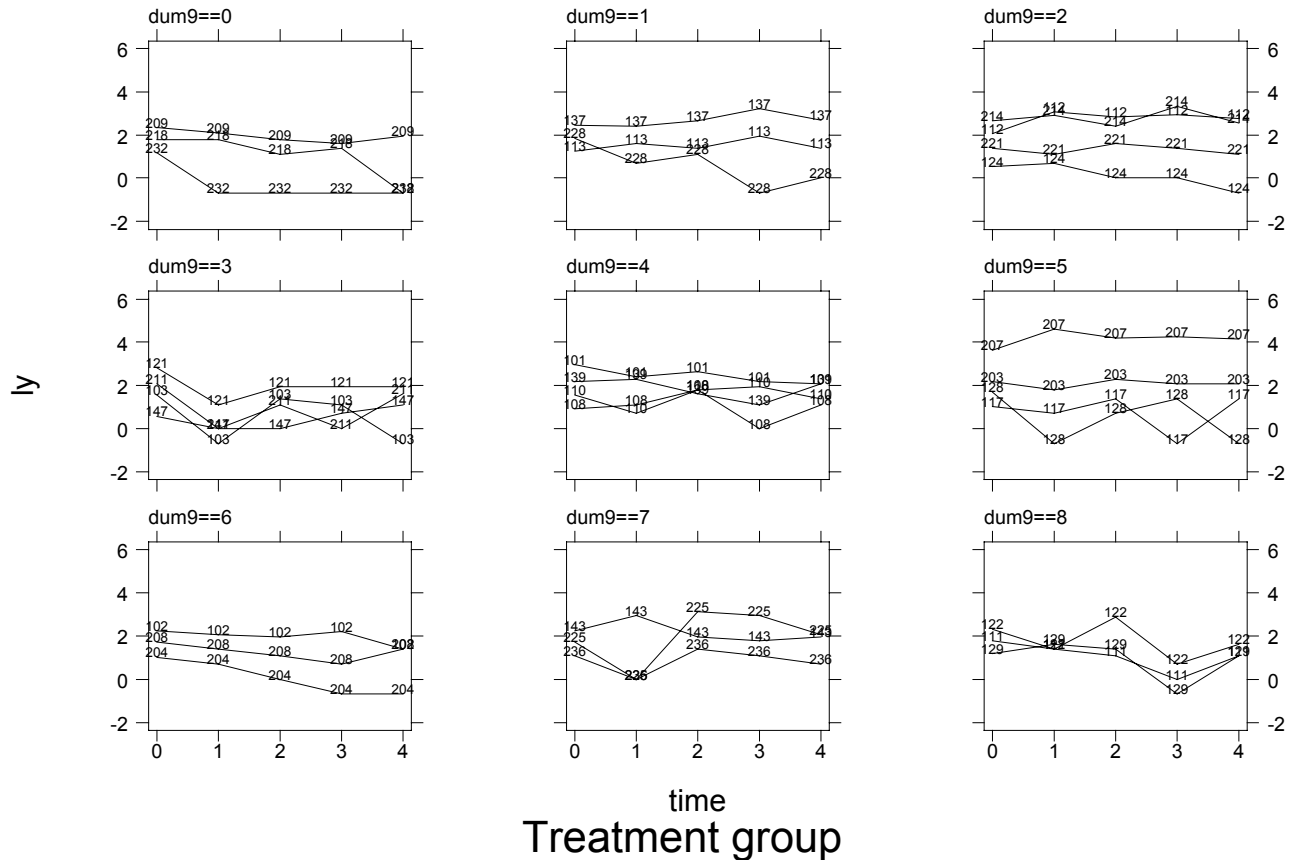

4

There is too much overlap between subjects; we use 9 graphs in each group to show the individual curve. First we produce a grouping variable dum9 that splits each treatment group into nine groups. If subjects are numbered by variable i, we can apply the function mod(i,9) to create a variable that numbers subjects from 0 to 8/ The graphs are clearer to read if within each treatment group and define i to number subjects in this order.

```
. sort treat baseline id time
. quietly by treat baseline id: gen i=_n==1
. replace i=sum(i)
(294 real changes made)
. gen dum9=mod(i,9)
. sort dum9 id time
. graph ly time if treat==0, c(L) s([id]) by(dum9) xlab ylab rlabel
b1("Placebo group")
```


Placebo group

```
. graph ly time if treat==1, c(L) s([id]) by(dum9) xlab ylab rlabel
b1("Treatment group")
```

Treatment group

Patient #207 had more epileptic seizures overall than any other subject. In the later regression analysis we will fit model with and without patient #207.

- **GEE marginal model**

$$\log E(Y_{ij}) = \log t_{ij} + \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} \quad i = 1,\ldots,59, \ j = 0,1,2,3,4$$

Where $t_{ij} = 8$ (weeks) if $j = 0$ and $t_{ij} = 2$ of $j = 1,2,3,4$. The $\log t_{ij}$ term is needed to account for different observation periods (offset). And the covariates are defined as:

$$x_{i1} = \begin{cases} 1 & \text{if visit 1, 2, 3 or 4} \\ 0 & \text{if baseline,} \end{cases} \qquad x_{i2} = \begin{cases} 1 & \text{if treatment group} \\ 0 & \text{if placebo group.} \end{cases}$$

```
. use   c:\data\seize, clear
. *rearrage the data to incorporate baseline data
. expand 2 if time==1
(59 observations created)
. sort id time
. quiet by id: replace time=0 if _n==1
. replace count =baseline if time==0
(59 real changes made)
. drop baseline

. *create the design matrix
. *offset time t
. gen t=log(2)
. replace t=log(8) if time==0
```

6

```
(59 real changes made)
. gen b1=1
. replace b1=0 if time==0
(59 real changes made)
. gen b2=treat
. gen b3=b1*b2

. *for complete data
. xtgee count b1 b2 b3, i(id) corr(exc) f(pois) l(log) offset(t) robust
scale(x2)

Iteration 1: tolerance = 7.828e-15

GEE population-averaged model              Number of obs     =       295
Group variable:                     id     Number of groups  =        59
Link:                              log     Obs per group: min =         5
Family:                        Poisson                    avg =       5.0
Correlation:              exchangeable                    max =         5
                                           Wald chi2(3)      =      0.90
Scale parameter:              19.43553     Prob > chi2       =    0.8251

                            (standard errors adjusted for clustering on id)
------------------------------------------------------------------------------
             |             Semi-robust
       count |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          b1 |   .1107981   .1170963     0.95   0.344    -.1187064    .3403027
          b2 |   .0265146   .2237583     0.12   0.906    -.4120436    .4650728
          b3 |  -.1036807   .2154436    -0.48   0.630    -.5259424    .3185811
       _cons |   1.347609   .1587079     8.49   0.000     1.036547    1.658671
           t |   (offset)
------------------------------------------------------------------------------
(Standard errors scaled using square root of Pearson X2-based dispersion)

. xtcorr

Estimated within-id correlation matrix R:

        c1       c2       c3       c4       c5
r1  1.0000
r2  0.7769   1.0000
r3  0.7769   0.7769   1.0000
r4  0.7769   0.7769   0.7769   1.0000
r5  0.7769   0.7769   0.7769   0.7769   1.0000

. *delete patient #207
. drop if id==207
(5 observations deleted)

. xtgee count b1 b2 b3, i(id) corr(exc) f(pois) l(log) offset(t) robust
scale(x2)

Iteration 1: tolerance = 9.506e-15

GEE population-averaged model              Number of obs     =       290
Group variable:                     id     Number of groups  =        58
Link:                              log     Obs per group: min =         5
Family:                        Poisson                    avg =       5.0
```

7

```
Correlation:                      exchangeable                    max =          5
                                                  Wald chi2(3)      =       3.67
Scale parameter:                      10.39709    Prob > chi2       =     0.2992

                                (standard errors adjusted for clustering on id)
-----------------------------------------------------------------------------
            |                 Semi-robust
      count |      Coef.    Std. Err.      z     P>|z|     [95% Conf. Interval]
-------------+---------------------------------------------------------------
         b1 |   .1107981    .1171137    0.95    0.344    -.1187405     .3403368
         b2 |   -.108028    .1953647   -0.55    0.580    -.4909357     .2748797
         b3 |  -.3015995    .1726956   -1.75    0.081    -.6400766     .0368777
       _cons |   1.347609    .1587315    8.49    0.000     1.036501    1.658717
          t |   (offset)
-----------------------------------------------------------------------------
(Standard errors scaled using square root of Pearson X2-based dispersion)

. xtcorr


Estimated within-id correlation matrix R:

          c1       c2       c3       c4       c5
r1   1.0000
r2   0.5983   1.0000
r3   0.5983   0.5983   1.0000
r4   0.5983   0.5983   0.5983   1.0000
r5   0.5983   0.5983   0.5983   0.5983   1.0000

. xtgee count b1 b2 b3, i(id) corr(exc) f(pois) l(log) offset(t) robust
scale(x2) eform


Iteration 1: tolerance = 9.506e-15


GEE population-averaged model            Number of obs      =        290
Group variable:                    id    Number of groups   =         58
Link:                             log    Obs per group: min =          5
Family:                       Poisson                   avg =        5.0
Correlation:              exchangeable                   max =          5
                                                  Wald chi2(3)      =       3.67
Scale parameter:                      10.39709    Prob > chi2       =     0.2992

                                (standard errors adjusted for clustering on id)
-----------------------------------------------------------------------------
            |                 Semi-robust
      count |        IRR    Std. Err.      z     P>|z|     [95% Conf. Interval]
-------------+---------------------------------------------------------------
         b1 |   1.117169    .1308359    0.95    0.344     .8880382    1.405421
         b2 |   .8976025    .1753598   -0.55    0.580     .6120534    1.316372
         b3 |   .7396343    .1277316   -1.75    0.081      .527252    1.037566
          t |   (offset)
-----------------------------------------------------------------------------
(Standard errors scaled using square root of Pearson X2-based dispersion)

. log close
```