```
xtreg count time, i(id)

Random-effects GLS regression                   Number of obs      =      2376
Group variable (i) : id                         Number of groups   =       369

R-sq:  within  = 0.2860                         Obs per group: min =         1
       between = 0.0895                                        avg =       6.4
       overall = 0.1750                                        max =        12

Random effects u_i ~ Gaussian                   Wald chi2(1)       =    833.71
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
   count |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    time |  -99.63042   3.450528    -28.874   0.000      -106.3933    -92.8675
   _cons |   836.9788   14.55013     57.524   0.000       808.4611    865.4966
---------+--------------------------------------------------------------------
 sigma_u |  252.15694
 sigma_e |  257.97343
     rho |  .48859953   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. xtgee count time, f(gaussian) corr(exc)

Iteration 1: tolerance = .11918583
Iteration 2: tolerance = .00056272
Iteration 3: tolerance = 3.294e-06
Iteration 4: tolerance = 1.932e-08

GEE population-averaged model                   Number of obs      =      2376
Group variable:                         id      Number of groups   =       369
Link:                             identity      Obs per group: min =         1
Family:                           Gaussian                     avg =       6.4
Correlation:                  exchangeable                     max =        12
                                                Wald chi2(1)       =    846.02
Scale parameter:                  132167.4      Prob > chi2        =    0.0000

------------------------------------------------------------------------------
   count |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    time |  -99.72951   3.428735    -29.086   0.000      -106.4497   -93.00931
   _cons |   836.9295   14.79482     56.569   0.000       807.9322    865.9268
------------------------------------------------------------------------------

. xtreg count time, i(id) re

Random-effects GLS regression                   Number of obs      =      2376
Group variable (i) : id                         Number of groups   =       369

R-sq:  within  = 0.2860                         Obs per group: min =         1
       between = 0.0895                                        avg =       6.4
       overall = 0.1750                                        max =        12

Random effects u_i ~ Gaussian                   Wald chi2(1)       =    833.71
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000
```

```
------------------------------------------------------------------------------
   count |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    time |  -99.63042   3.450528    -28.874   0.000      -106.3933    -92.8675
   _cons |   836.9788   14.55013     57.524   0.000       808.4611    865.4966
---------+--------------------------------------------------------------------
 sigma_u |  252.15694
 sigma_e |  257.97343
     rho |  .48859953   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. clear

. use "A:\logistic.dta", clear

. * consider the following data for logistic regression
. glm y, f(bin) l(logit)

Iteration 1 : deviance =   34.4535
Iteration 2 : deviance =   34.3718
Iteration 3 : deviance =   34.3718
Iteration 4 : deviance =   34.3718

Residual df  =        26                              No. of obs =        27
Pearson X2   =  26.99999                              Deviance   =  34.37177
Dispersion   =  1.038461                              Dispersion =  1.321991

Bernoulli distribution, logit link
------------------------------------------------------------------------------
       y |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   _cons |  -.6931472   .4082482     -1.698   0.090      -1.493299    .1070046
------------------------------------------------------------------------------

. * look at the deviance its 34.37
. * now lets include li in the model
. glm y li, f(bin) l(logit)

Iteration 1 : deviance =   26.1073
Iteration 2 : deviance =   26.0730
Iteration 3 : deviance =   26.0730
Iteration 4 : deviance =   26.0730

Residual df  =        25                              No. of obs =        27
Pearson X2   =  23.93291                              Deviance   =  26.07296
Dispersion   =  .9573164                              Dispersion =  1.042919

Bernoulli distribution, logit link
------------------------------------------------------------------------------
       y |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      li |   2.897264    1.18682      2.441   0.015       .5711401    5.223387
   _cons |   -3.77714   1.378624     -2.740   0.006      -6.479194   -1.075087
------------------------------------------------------------------------------

. * the deviance decresed to 26.07, also li is significant
. * now lets fit a model with a constant term and temp
```

```
. glm y temp, f(bin) l(logit)

Iteration 1 : deviance =   33.8750
Iteration 2 : deviance =   33.8180
Iteration 3 : deviance =   33.8180
Iteration 4 : deviance =   33.8180

Residual df  =        25                        No. of obs =        27
Pearson X2   =   26.8173                        Deviance   =  33.81799
Dispersion   =  1.072692                        Dispersion =  1.352719

Bernoulli distribution, logit link
------------------------------------------------------------------------------
       y |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    temp |  -22.01816   30.76225     -0.716   0.474      -82.31106    38.27475
   _cons |   21.23275   30.60836      0.694   0.488      -38.75853    81.22403
------------------------------------------------------------------------------

. * not much decrease in deviance. Also the coefficiant is not significant. Thu
> s temp dosent seem to be important to explain the variation in the response
. * now lets include both li and temp in the model
. glm y li temp, f(bin) l(logit)

Iteration 1 : deviance =   25.3826
Iteration 2 : deviance =   24.8337
Iteration 3 : deviance =   24.7970
Iteration 4 : deviance =   24.7968
Iteration 5 : deviance =   24.7968

Residual df  =        24                        No. of obs =        27
Pearson X2   =  21.83892                        Deviance   =  24.79676
Dispersion   =  .9099551                        Dispersion =  1.033198

Bernoulli distribution, logit link
------------------------------------------------------------------------------
       y |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      li |   3.298405   1.364846      2.417   0.016       .6233552    5.973454
    temp |  -49.98084   47.91913     -1.043   0.297      -143.9006    43.93892
   _cons |   45.40973   46.83838      0.969   0.332      -46.39181    137.2113
------------------------------------------------------------------------------

. * from the above results it is seen that temp is not significant alone or wit
> h li. Thus we decide that the final model will include only one covariate li
. glm y li, f(bin) l(logit)

Iteration 1 : deviance =   26.1073
Iteration 2 : deviance =   26.0730
Iteration 3 : deviance =   26.0730
Iteration 4 : deviance =   26.0730

Residual df  =        25                        No. of obs =        27
Pearson X2   =  23.93291                        Deviance   =  26.07296
Dispersion   =  .9573164                        Dispersion =  1.042919

Bernoulli distribution, logit link
```

```
------------------------------------------------------------------------------
       y |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      li |   2.897264    1.18682       2.441   0.015       .5711401    5.223387
   _cons |   -3.77714   1.378624      -2.740   0.006      -6.479194   -1.075087
------------------------------------------------------------------------------

. * thus we can write the model as logit(p) = -3.778 + 2.897 li
. * if one wishes to obtain the estimate of odds ratio you can use either of th
> e following commands
. disp exp(2.897264)
18.124489

. glm y li, f(bin) l(logit) eform

Iteration 1 : deviance =   26.1073
Iteration 2 : deviance =   26.0730
Iteration 3 : deviance =   26.0730
Iteration 4 : deviance =   26.0730

Residual df  =        25                              No. of obs =        27
Pearson X2   =  23.93291                              Deviance   =  26.07296
Dispersion   =  .9573164                              Dispersion =  1.042919

Bernoulli distribution, logit link
------------------------------------------------------------------------------
       y | Odds Ratio   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      li |   18.12448   21.51049       2.441   0.015       1.770284    185.5617
------------------------------------------------------------------------------

. * we obtaint the OR as 18.13
. clear

. edit
- preserve

. * Now we demonstrate to do a logistic regression analysis when we have the da
> ta in the for Bin(n p)
. glm blind, f(bin 50) l(logit)

Iteration 1 : deviance =  106.1028
Iteration 2 : deviance =  105.7517
Iteration 3 : deviance =  105.7517

Residual df  =         4                              No. of obs =         5
Pearson X2   =  95.03189                              Deviance   =  105.7517
Dispersion   =  23.75797                              Dispersion =  26.43794

Binomial (N=50) distribution, logit link
------------------------------------------------------------------------------
   blind |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   _cons |  -.0800427   .1265923      -0.632   0.527      -.3281591    .1680737
------------------------------------------------------------------------------

. * look at the deviance
```

```
. glm blind  age, f(bin 50) l(logit)

Iteration 1 : deviance =    6.6044
Iteration 2 : deviance =    6.4473
Iteration 3 : deviance =    6.4471
Iteration 4 : deviance =    6.4471

Residual df  =         3                              No. of obs =         5
Pearson X2   =   6.13217                              Deviance   =  6.447147
Dispersion   =  2.044057                              Dispersion =  2.149049

Binomial (N=50) distribution, logit link
------------------------------------------------------------------------------
   blind |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |   .0940683   .0119755      7.855   0.000       .0705967    .1175399
   _cons |  -4.356181   .5700965     -7.641   0.000      -5.473549   -3.238812
------------------------------------------------------------------------------

. * the deviance reduced to 6.44 after we include the covariate in the model. A
> lso the age coefficient is highly significant. Thus the final model will be o
> f the form logit(p) = -4.36 + 0.094 Age. After the glm commant one has an opt
> ion to give a commant to predict the fitted values. One can also obtain the o
> dds ratio by using the option eform in the glm command
. predict fit
(option mu assumed; predicted mean blind)

. * now we look at the corresponding observed and estimated probabilities
. gen op = bilnd/50
bilnd not found
r(111);

. gen op = blind/50

. gen fp =  fit/50

. *not lets grag the observed and the fitted probabilities
. garph op fp age
unrecognized command:  garph
r(199);

. graph op fp age

. * to obatin the OR
. glm blind  age, f(bin 50) l(logit) eform

Iteration 1 : deviance =    6.6044
Iteration 2 : deviance =    6.4473
Iteration 3 : deviance =    6.4471
Iteration 4 : deviance =    6.4471

Residual df  =         3                              No. of obs =         5
Pearson X2   =   6.13217                              Deviance   =  6.447147
Dispersion   =  2.044057                              Dispersion =  2.149049

Binomial (N=50) distribution, logit link
------------------------------------------------------------------------------
```

```
   blind | Odds Ratio   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |   1.098635   .0131567      7.855   0.000       1.073148    1.124726
------------------------------------------------------------------------------

. * the corresponding OR is 1.099
. clear

. close
unrecognized command:  close
r(199);

.
```