Parametric analysis (II part)

Data description: The "ovarian" dataset contains information on survival in a randomized trial comparing two treatments group for ovarian cancer

1. Working on the pooled sample:

```
. cd "working directory path"
. use ovarian.dta,clear
. des
Contains data from ovarian.dta
  obs:            26
 vars:             7                         7 May 2015 15:22
 size:           390 (99.9% of memory free)
-----------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-----------------------------------------------------------------------------------
id              byte   %8.0g                  identity number
futime          int    %8.0g                  survival or censoring time
fustat          byte   %8.0g                  censoring status
age             float  %9.0g                  age (in years)
resids          byte   %8.0g                  residual disease present (1=no, 2=yes)
rx              byte   %8.0g                  treatment group
ecogps          byte   %8.0g                  ECOG performance status (1 if better)
-----------------------------------------------------------------------------------
. recode rx (2=0)
```

2. **Preliminary analysis** Describe data looking at frequency distributions of variables and at thier cross-tabulations

```
. tab rx
  treatment |
      group |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |         13       50.00       50.00
          2 |         13       50.00      100.00
------------+-----------------------------------
      Total |         26      100.00

. tab residds
   residual |
    disease |
    present |
     (1=no, |
     2=yes) |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |         11       42.31       42.31
          2 |         15       57.69      100.00
------------+-----------------------------------
      Total |         26      100.00
```

```
. tab ecogps
      ECOG |
performance |
  status (1 |
 if better) |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |         14       53.85       53.85
          2 |         12       46.15      100.00
------------+-----------------------------------
      Total |         26      100.00

. sum age,d
                       age (in years)
-------------------------------------------------------------
      Percentiles      Smallest
 1%      38.8932        38.8932
 5%      39.2712        39.2712
10%      43.1233        43.1233       Obs                  26
25%      50.1096         43.137       Sum of Wgt.          26

50%      56.8466                      Mean           56.16544
                        Largest       Std. Dev.      10.10036
75%      63.2192        66.4658
90%      72.3315        72.3315       Variance       102.0173
95%      74.4932        74.4932       Skewness       .0434026
99%      74.5041        74.5041       Kurtosis       2.343786

. ttest age,by(rx)
Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       1 |      13    55.73235    3.749421    13.51873    47.56307    63.90164
       2 |      13    56.59853    1.503165    5.419739    53.32342    59.87365
---------+--------------------------------------------------------------------
combined |      26    56.16544    1.980844    10.10036    52.08582    60.24507
---------+--------------------------------------------------------------------
    diff |           -.8661781    4.039513               -9.203323    7.470966
------------------------------------------------------------------------------
    diff = mean(1) - mean(2)                                      t =  -0.2144
Ho: diff = 0                                     degrees of freedom =       24

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.4160         Pr(|T| > |t|) = 0.8320          Pr(T > t) = 0.5840

. tab residds rx,col row chi2
  residual |
   disease |
```

```
   present |
   (1=no, |    treatment group
   2=yes) |         1          2 |     Total
-----------+----------------------+----------
         1 |         5          6 |        11
           |     45.45      54.55 |    100.00
           |     38.46      46.15 |     42.31
-----------+----------------------+----------
         2 |         8          7 |        15
           |     53.33      46.67 |    100.00
           |     61.54      53.85 |     57.69
-----------+----------------------+----------
     Total |        13         13 |        26
           |     50.00      50.00 |    100.00
           |    100.00     100.00 |    100.00

          Pearson chi2(1) =   0.1576   Pr = 0.691

. tab ecogps rx, col row chi2
      ECOG |
performanc |
  e status |
     (1 if |    treatment group
   better) |         1          2 |     Total
-----------+----------------------+----------
         1 |         7          7 |        14
           |     50.00      50.00 |    100.00
           |     53.85      53.85 |     53.85
-----------+----------------------+----------
         2 |         6          6 |        12
           |     50.00      50.00 |    100.00
           |     46.15      46.15 |     46.15
-----------+----------------------+----------
     Total |        13         13 |        26
           |     50.00      50.00 |    100.00
           |    100.00     100.00 |    100.00

          Pearson chi2(1) =   0.0000   Pr = 1.000

. tab residds ecogps, col row chi2
  residual |
   disease |
   present |   ECOG performance
   (1=no, | status (1 if better)
   2=yes) |         1          2 |     Total
-----------+----------------------+----------
         1 |         5          6 |        11
           |     45.45      54.55 |    100.00
           |     35.71      50.00 |     42.31
```

```
-----------+----------------------+----------
         2 |         9          6 |        15 
           |     60.00      40.00 |    100.00 
           |     64.29      50.00 |     57.69 
-----------+----------------------+----------
     Total |        14         12 |        26 
           |     53.85      46.15 |    100.00 
           |    100.00     100.00 |    100.00 

          Pearson chi2(1) =   0.5403   Pr = 0.462

. ttest age,by(residds)
Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       1 |      11     52.7721    2.280056    7.562091    47.69182    57.85238
       2 |      15    58.65389    2.894191    11.20915    52.44647    64.86132
---------+--------------------------------------------------------------------
combined |      26    56.16544    1.980844    10.10036    52.08582    60.24507
---------+--------------------------------------------------------------------
    diff |           -5.881793    3.912005               -13.95577    2.192188
------------------------------------------------------------------------------
    diff = mean(1) - mean(2)                                      t =  -1.5035
Ho: diff = 0                                     degrees of freedom =       24

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0729         Pr(|T| > |t|) = 0.1458          Pr(T > t) = 0.9271

. ttest age,by(ecogps)
Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       1 |      14    54.98004    2.900216    10.85161    48.71451    61.24558
       2 |      12    57.54841    2.720518    9.424152    51.56059    63.53623
---------+--------------------------------------------------------------------
combined |      26    56.16544    1.980844    10.10036    52.08582    60.24507
---------+--------------------------------------------------------------------
    diff |           -2.568365    4.021366               -10.86806    5.731326
------------------------------------------------------------------------------
    diff = mean(1) - mean(2)                                      t =  -0.6387
Ho: diff = 0                                     degrees of freedom =       24

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.2645         Pr(|T| > |t|) = 0.5291          Pr(T > t) = 0.7355
```

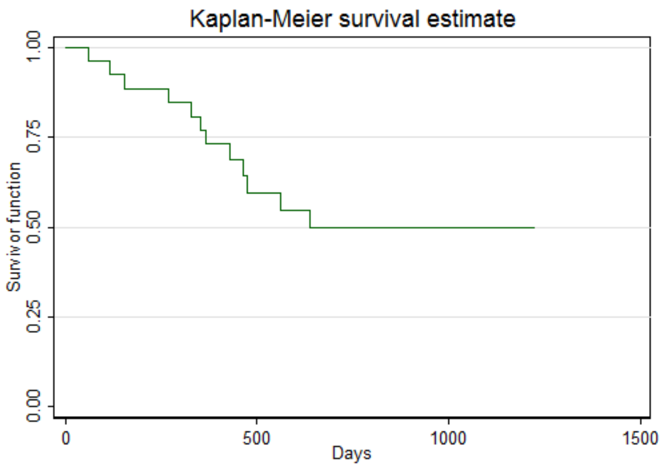
3. Declare data to be survival-data in order to take into account the follow-up

```
. stset futime,fail(fustat)id(id)
```

```
                 id:  id
      failure event:  fustat != 0 & fustat < .
obs. time interval:  (futime[_n-1], futime]
 exit on or before:  failure

------------------------------------------------------------------------------
       26  total obs.
        0  exclusions
------------------------------------------------------------------------------
       26  obs. remaining, representing
       26  subjects
       12  failures in single failure-per-subject data
    15588  total analysis time at risk, at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =      1227
```

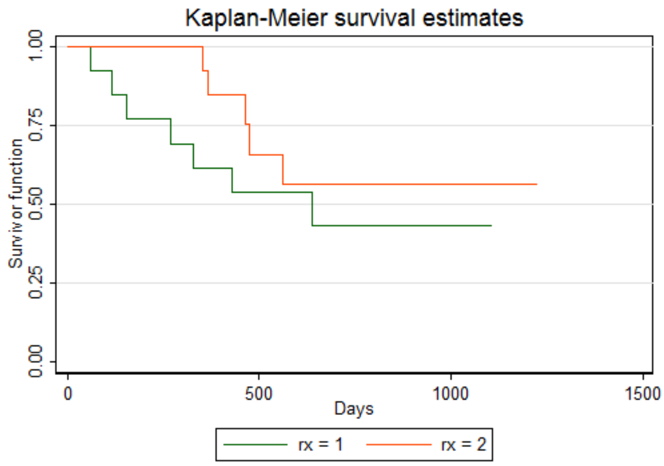
4. By using non-parametric methods represent the total survival curve and the survival curves by variables by Kaplan-Meier estimator and test their difference

```
. sts graph
```

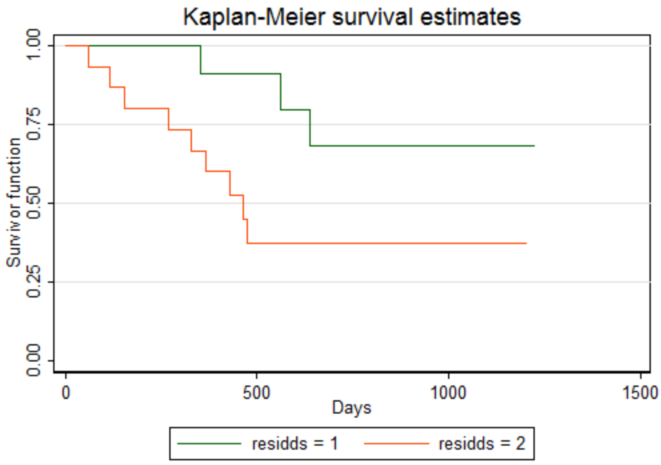



```
. sts graph,by(rx)
. sts test rx
Log-rank test for equality of survivor functions

      |   Events         Events
rx    |  observed       expected
------+-------------------------
1     |         7           5.23
2     |         5           6.77
------+-------------------------
```

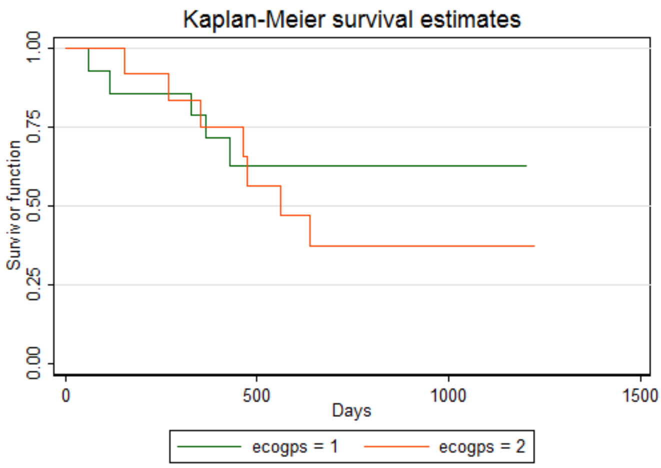
```
Total |         12          12.00

            chi2(1) =       1.06
            Pr>chi2 =     0.3026
```




```
. sts graph,by(residds)
. sts test residds
Log-rank test for equality of survivor functions

        |  Events          Events
residds |  observed        expected
--------+-------------------------
1       |         3            6.26
2       |         9            5.74
--------+-------------------------
Total   |        12           12.00

              chi2(1) =        3.62
              Pr>chi2 =      0.0570

. sts graph,by(ecogps)
. sts test ecogps
Log-rank test for equality of survivor functions

       |  Events          Events
ecogps |  observed        expected
-------+-------------------------
1      |         5            6.18
2      |         7            5.82
-------+-------------------------
Total  |        12           12.00
```

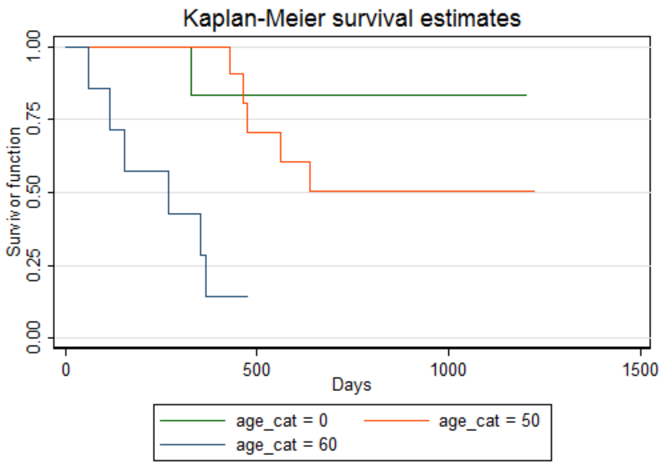



```
chi2(1) =        0.47
Pr>chi2 =      0.4941
```



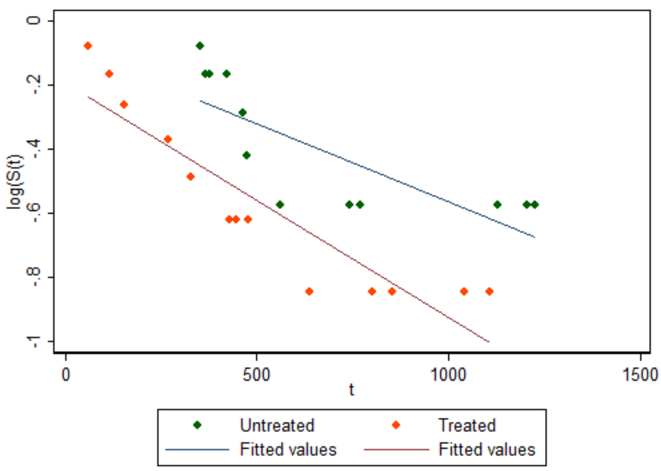

```
. egen age_cat=cut(age),at(0,50(10)60,100)
. sts graph,by(age_cat)
. sts test age_cat
Log-rank test for equality of survivor functions

        |  Events          Events
age_cat |  observed        expected
--------+------------------------
0       |         1            3.60
50      |         5            6.93
60      |         6            1.47
```

```
--------+------------------------
Total   |         12         12.00

              chi2(2) =      17.68
              Pr>chi2 =     0.0001
```

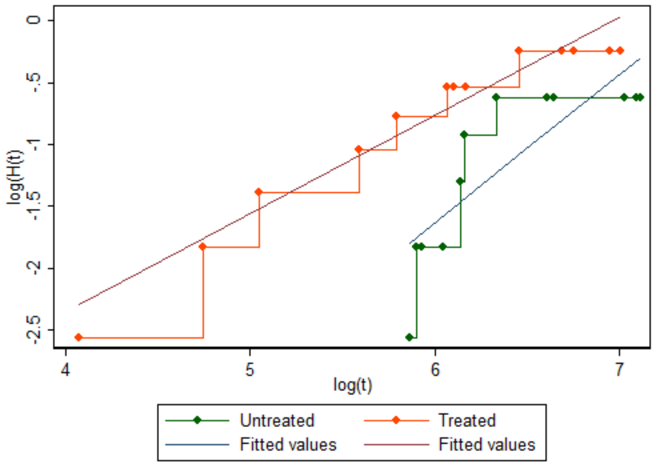


5. Before fitting the regression models, check for the assumptions underlying the exponential, weibull, log-logistic and log-normal models respectively by graphical methods

```
. *exponential model
. *survival function
. sts gen km= s,by(rx)
. gen logkm_exp=log(km)
. twoway (scatter logkm_exp _t if rx==0,sort)(scatter logkm_exp _t if rx==1,sort) ///
> (lfit logkm_exp _t if rx==0,sort) (lfit logkm_exp _t if rx==1,sort), ///
> legend(lab(1 "Untreated") lab(2 "Treated")) ytitle("log(S(t)") xtitle("t")
```

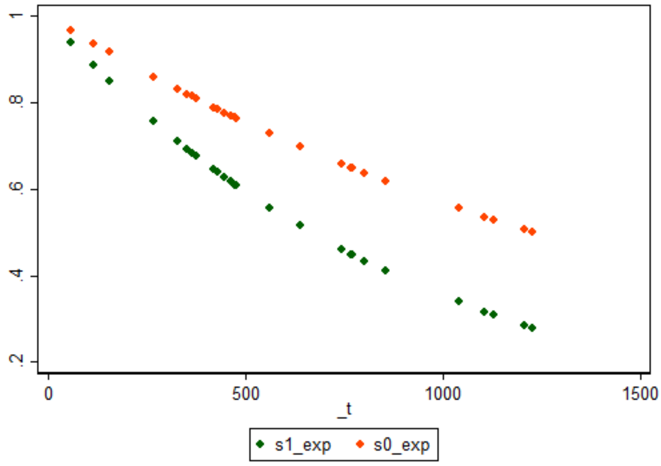


```
. *cumulative hazard function
. sts gen cumhaz_exp=na,by(rx)
. gen cumhaz_exp1=cumhaz_exp
. gen cumhaz_exp0=cumhaz_exp
. gen logcumhaz_exp=log(cumhaz_exp)
. gen logtime=log(_t)
. twoway (scatter logcumhaz_exp logtime if rx==0,connect(J)sort) ///
> (scatter logcumhaz_exp logtime if rx==1,connect(J)sort) ///
> (lfit logcumhaz_exp logtime if rx==0) (lfit logcumhaz_exp logtime if rx==1), ///
> legend(lab(1 "Untreated") lab(2 "Treated")) ytitle("log(H(t)") xtitle("log(t)")
```

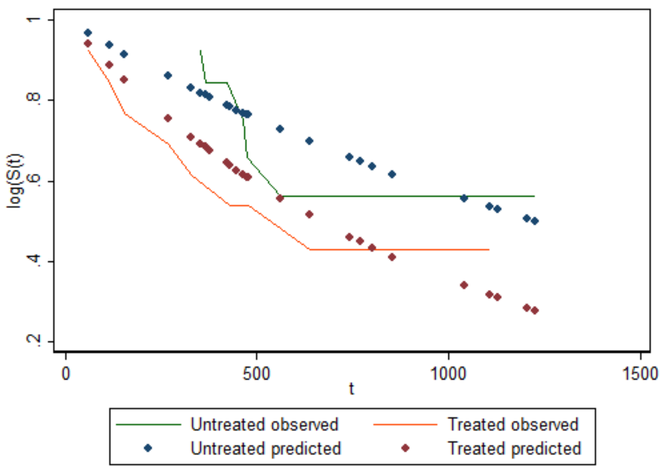



```
. *observed vs predicted
. xi:streg rx,dist(e) nohr
------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          rx |   .6125256     .58554     1.05   0.296    -.5351118    1.760163
       _cons |  -7.480203   .4472136   -16.73   0.000    -8.356725    -6.60368
------------------------------------------------------------------------------
*observed vs predicted
. xi:streg rx,dist(e) nohr
. predict surv_exp,surv
. twoway (line km _t if rx==0,sort)(line km _t if rx==1,sort) ///
> (scatter surv_exp _t if rx==0,sort) (scatter surv_exp _t if rx==1,sort), ///
> legend(lab(1 "Untreated observed") lab(2 "Treated observed") lab(3 "Untreated predicted")
> lab(4 "Treated predicted")) ytitle("log(S(t)") xtitle("t")

*alternatively
. gen s1_exp=exp(-exp(_b[_cons]+_b[rx])*_t)
. gen s0_exp=exp(-exp(_b[_cons])*_t)
. sort _t
. twoway (scatter s1_exp _t) (scatter s0_exp _t)
```

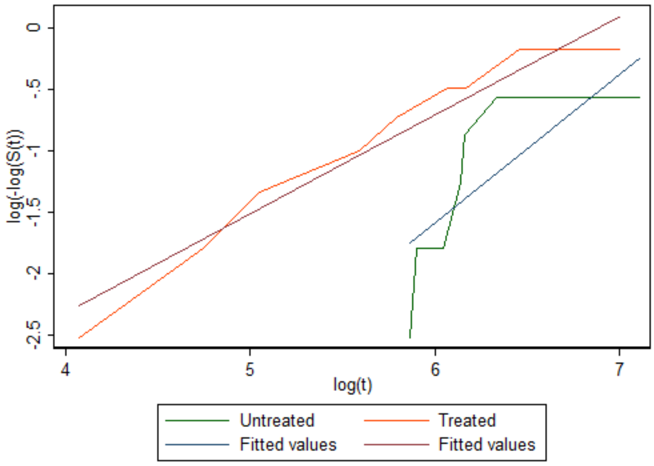
```
. stcurve, surv at(rx=0) at1(rx=1)
. twoway (line km _t if rx==0,sort)(line km _t if rx==1,sort) ///
> (scatter s0_exp _t,sort) (scatter s1_exp _t,sort), ///
> legend(lab(1 "Untreated observed") lab(2 "Treated observed") lab(3 "Untreated predicted")
```



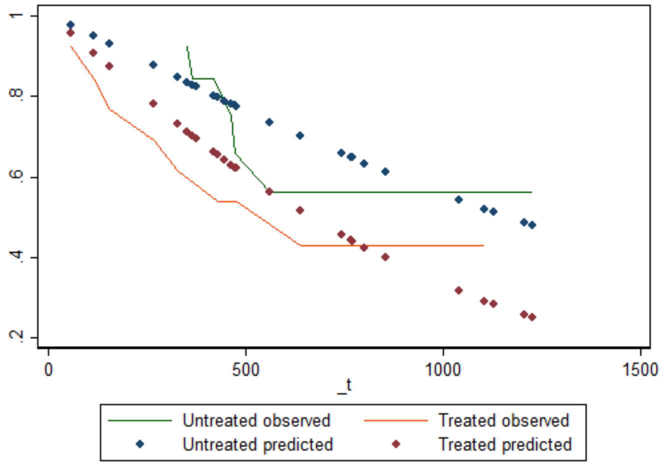

```
. twoway (line km _t if rx==0,sort)(line km _t if rx==1,sort) ///
> (scatter s0_exp _t,sort) (scatter s1_exp _t,sort), legend(lab(1 "Untreated observed") ///
> lab(2 "Treated observed") lab(3 "Untreated predicted") lab(4 "Treated predicted")) ytitle(
```

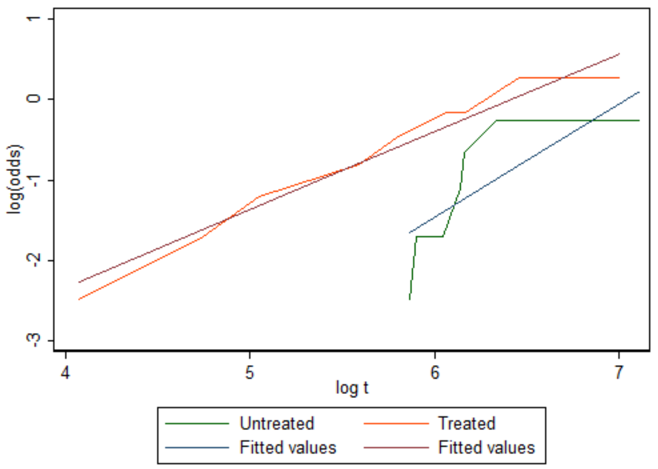



```
. *weibull model
. *survival function
. gen logkm_w=log(-log(km))
. gen logt=log(_t)
. twoway (line logkm_w logt if rx==0,sort)(line logkm_w logt if rx==1,sort) ///
> (lfit logkm_w logt if rx==0,sort) (lfit logkm_w logt if rx==1,sort), ///
> legend(lab(1 "Untreated") lab(2 "Treated")) ytitle("log(-log(S(t))") xtitle("log(t)")
```

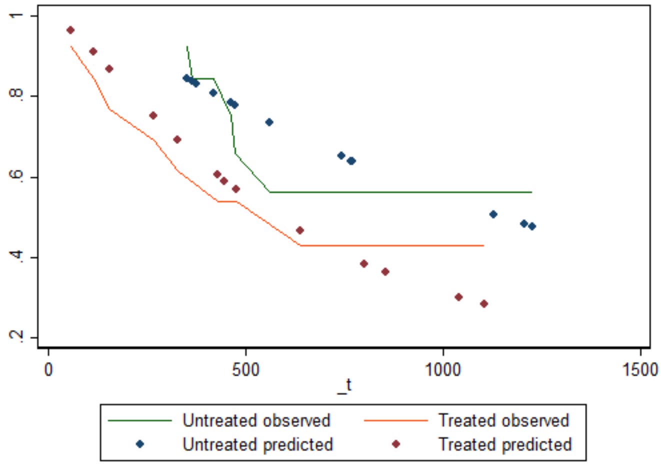



```
. *observed vs predicted
. xi:streg rx,dist(w) nohr
------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          rx |   .6314782   .5868807     1.08   0.282     -.518787    1.781743
       _cons |  -8.337065   1.940367    -4.30   0.000    -12.14011   -4.534017
-------------+----------------------------------------------------------------
       /ln_p |   .1214247   .2514893     0.48   0.629    -.3714853    .6143346
-------------+----------------------------------------------------------------
           p |   1.129104   .2839576                      .6897092    1.848426
         1/p |   .8856578   .2227334                      .5410008    1.449886
------------------------------------------------------------------------------
. gen s_w0=exp(-exp(_b[_cons])*_t^1.129104)
. gen s_w1=exp(-exp(_b[_cons]+_b[rx])*_t^1.129104)
. twoway (scatter s_w0 _t) (scatter s_w1 _t)
. stcurve, surv at(rx=0) at1(rx=1)
. twoway (line km _t if rx==0,sort)(line km _t if rx==1,sort) ///
> (scatter s_w0 _t,sort) (scatter s_w1 _t,sort), ///
> legend(lab(1 "Untreated observed") lab(2 "Treated observed") ///
> lab(3 "Untreated predicted") lab(4 "Treated predicted"))


. *log-logistic model
. gen odds_km=(1-km)/km
. gen logodds=log(odds_km)
. twoway (line logodds logt if rx==0,sort)(line logodds logt if rx==1,sort) ///
> (lfit logodds logt if rx==0,sort) (lfit logodds logt if rx==1,sort), ///
> legend(lab(1 "Untreated") lab(2 "Treated")) ytitle("log(odds)") xtitle("log t")


. *observed vs predicted
. xi:streg rx,dist(loglog)
------------------------------------------------------------------------------
```

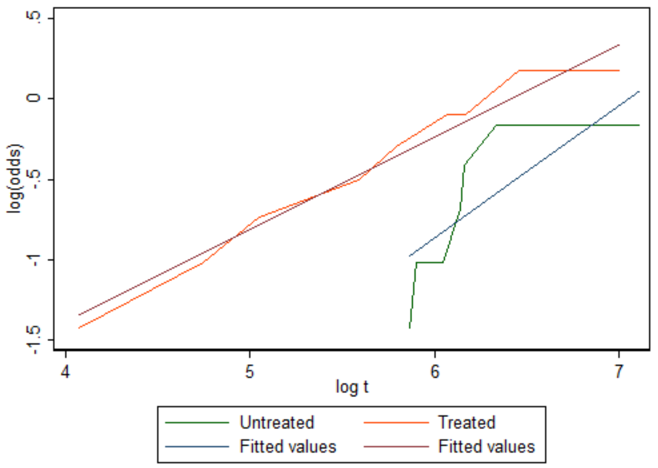





```
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          rx |  -.6851239   .5365499    -1.28   0.202    -1.736742    .3664945
       _cons |   7.046231   .4118578    17.11   0.000     6.239005    7.853458
-------------+----------------------------------------------------------------
     /ln_gam |  -.3594936   .2466853    -1.46   0.145    -.8429879    .1240007
-------------+----------------------------------------------------------------
       gamma |   .6980297   .1721937                      .4304225    1.132017
------------------------------------------------------------------------------
. predict surv_loglog,surv
. twoway (line km _t if rx==0,sort)(line km _t if rx==1,sort) ///
> (scatter surv_loglog _t if rx==0,sort) (scatter surv_loglog _t if rx==1,sort), ///
> legend(lab(1 "Untreated observed") lab(2 "Treated observed") ///
lab(3 "Untreated predicted") lab(4 "Treated predicted"))


. *log-normal model
```

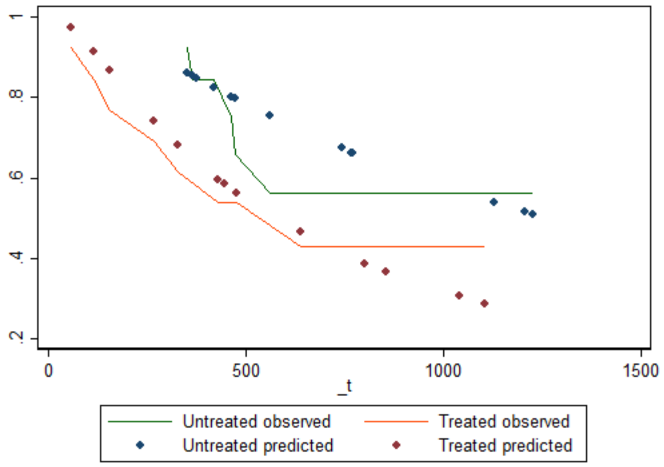



```
. gen inv_km=invnormal(1-km)
. twoway (line inv_km logt if rx==0,sort)(line inv_km logt if rx==1,sort) ///
> (lfit inv_km logt if rx==0,sort) (lfit inv_km logt if rx==1,sort), ///
> legend(lab(1 "Untreated") lab(2 "Treated")) ytitle("log(odds)") xtitle("log t")
```




```
. *observed vs predicted
. xi:streg rx,dist(ln)
------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          rx |  -.7864423   .5269101    -1.49   0.136    -1.819167    .2462826
       _cons |   7.140936   .4239738    16.84   0.000     6.309963    7.971909
-------------+----------------------------------------------------------------
     /ln_sig |   .1600969   .2267412     0.71   0.480    -.2843076    .6045015
-------------+----------------------------------------------------------------
       sigma |   1.173625    .266109                      .7525351    1.830339
```

```
--------------------------------------------------------------------------------
. predict surv_ln,surv
```



6. By using the exponential model evaluate the treatment effect (crude and adjusted estimates) on mortality

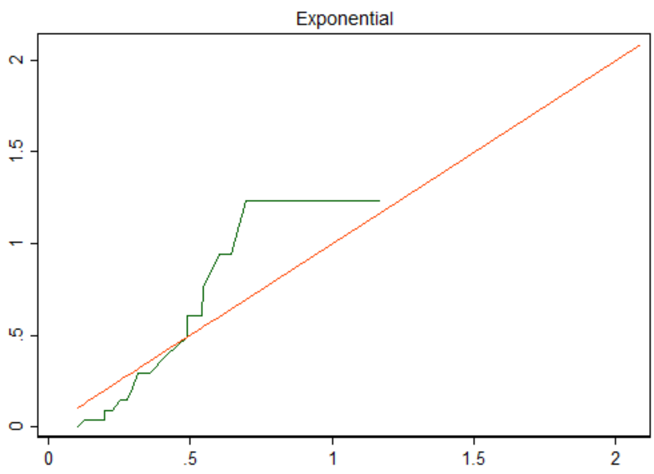
```
. xi:streg i.rx,dist(e)
Log likelihood  =   -29.288092                     Prob > chi2     =    0.2910

--------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+------------------------------------------------------------------
      _Irx_1 |   1.845086   1.080371     1.05   0.296     .5856038    5.813385
--------------------------------------------------------------------------------

. xi:streg i.rx age,dist(e)
Log likelihood  =   -23.022327                     Prob > chi2     =    0.0011

--------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+------------------------------------------------------------------
      _Irx_1 |    1.93685   1.192845     1.07   0.283     .5792526    6.476258
         age |   1.110781   .0354083     3.30   0.001     1.043506    1.182394
--------------------------------------------------------------------------------

. xi:streg i.rx i.residds age  i.ecogps,dist(e)
Log likelihood  =   -22.457651                     Prob > chi2     =    0.0052

--------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+------------------------------------------------------------------
      _Irx_1 |    1.87184   1.153513     1.02   0.309     .5593961    6.263511
 _Iresidds_2 |    2.15096   1.594008     1.03   0.301     .5033083    9.192438
```

```
         age |   1.091417   .0368664     2.59   0.010       1.0215    1.166119
  _Iecogps_2 |   1.287025   .7800968     0.42   0.677     .3923316    4.222024
------------------------------------------------------------------------------

. xi:streg i.rx*i.ecogps i.residds age,dist(e)
Log likelihood  =   -21.612988                     Prob > chi2     =    0.0056

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      _Irx_1 |   5.318805   6.053983     1.47   0.142     .5714224    49.50748
  _Iecogps_2 |   3.638872   4.113217     1.14   0.253     .3970149    33.35238
_IrxXeco_1_2 |   .1914859   .2593592    -1.22   0.222     .0134655    2.723027
 _Iresidds_2 |   2.135929   1.589535     1.02   0.308     .4967432    9.184209
         age |   1.090369    .037685     2.50   0.012     1.018954     1.16679
------------------------------------------------------------------------------

. xi:streg i.rx*i.residds age  i.ecogps,dist(e)
Log likelihood  =   -22.321705                     Prob > chi2     =    0.0102

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      _Irx_1 |   1.085295   1.355736     0.07   0.948     .0938076    12.55619
 _Iresidds_2 |   1.595025   1.460327     0.51   0.610      .265124    9.595908
_IrxXres_1_2 |   2.179131   3.296521     0.51   0.607     .1123602    42.26241
         age |   1.083755   .0372925     2.34   0.019     1.013074    1.159369
  _Iecogps_2 |   1.339869   .8126651     0.48   0.630     .4081191    4.398838
------------------------------------------------------------------------------

. xi:streg i.rx i.residds age  i.ecogps,dist(e)
Log likelihood  =   -22.457651                     Prob > chi2     =    0.0052

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      _Irx_1 |    1.87184   1.153513     1.02   0.309     .5593961    6.263511
 _Iresidds_2 |    2.15096   1.594008     1.03   0.301     .5033083    9.192438
         age |   1.091417   .0368664     2.59   0.010       1.0215    1.166119
  _Iecogps_2 |   1.287025   .7800968     0.42   0.677     .3923316    4.222024
------------------------------------------------------------------------------
. est store a
. xi:streg i.rx i.residds i.age_cat  i.ecogps,dist(e)
Log likelihood  =   -22.151354                     Prob > chi2     =    0.0088

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      _Irx_1 |   1.866375   1.312079     0.89   0.375     .4705396    7.402894
```

```
 _Iresidds_2 |     2.3475   1.724604     1.16   0.245     .5562446    9.907075
_Iage_cat_50 |   4.503212   5.283785     1.28   0.200     .4516232     44.9023
_Iage_cat_60 |   14.74248     16.171     2.45   0.014     1.717469    126.5471
  _Iecogps_2 |   1.773817   1.067332     0.95   0.341     .5454163    5.768852
------------------------------------------------------------------------------
. est store b
. lrtest a b
Likelihood-ratio test                                 LR chi2(1)  =      0.61
(Assumption: a nested in b)                           Prob > chi2 =    0.4338
. xi:streg i.rx i.residds age  i.ecogps,dist(e)
Log likelihood  =   -22.457651                    Prob > chi2     =    0.0052

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _Irx_1 |    1.87184   1.153513     1.02   0.309     .5593961    6.263511
 _Iresidds_2 |    2.15096   1.594008     1.03   0.301     .5033083    9.192438
         age |   1.091417   .0368664     2.59   0.010       1.0215    1.166119
  _Iecogps_2 |   1.287025   .7800968     0.42   0.677     .3923316    4.222024
------------------------------------------------------------------------------
```

7. Calculate Cox-Snell residuals for the exponential model (post-estimation)

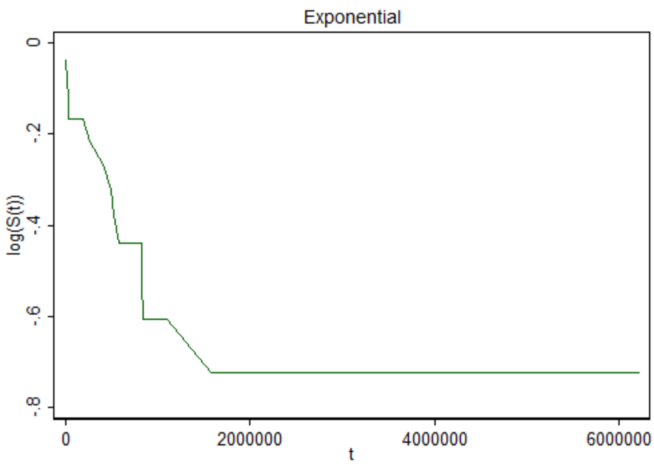
```
. predict residual, csnell
. stset residual, failure(fustat)id(id)
. sts gen km = s
. generate cumhaz = -ln(km)
. line cumhaz residual residual, sort name("exponential, replace") ///
> legend(off) t1(Exponential) xtitle("")
. drop cumhaz residual k
```



8. Calculate standardized residuals for the exponential model (post-estimation)

```
. stset futime,fail(fustat)id(id)
```

```
. xi:streg i.rx i.residds age  i.ecogps,dist(e)
. predict pred, xb
. gen res_stand=(log(_t)-pred)
. replace res_stand=exp(res_stand)
. stset res_stand, failure(fustat)id(id)
. sts gen km = s
. gen logkm=log(km)
. line logkm _t,sort legend(off) t1(Exponential) xtitle("t") ytitle("log(S(t))")
. drop km logkm res_stand pred
```



9. Calculate deviance residuals for the exponential model (post-estimation)

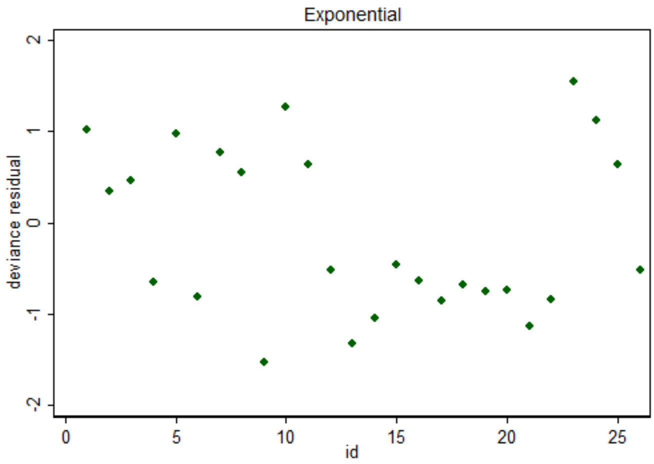
```
. stset futime,fail(fustat)id(id)
. xi:streg i.rx i.residds age  i.ecogps,dist(e)
. predict deviance, deviance
. graph twoway scatter deviance id, legend(off) t1(Exponential) xtitle("id")
. drop deviance
```

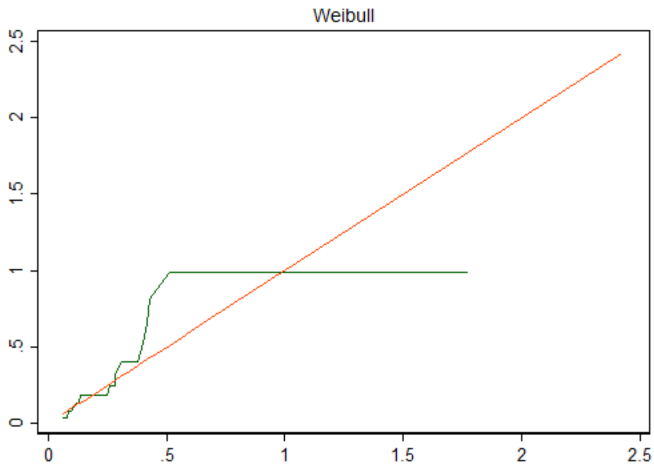
10. After fitting the Weibull model, calculate Cox-Snell residuals (post-estimation)

```
. stset futime,fail(fustat)id(id)
. xi:streg i.rx i.residds age  i.ecogps,dist(w)
. predict residual, csnell
. stset residual, failure(fustat)id(id)
. sts gen km = s
. generate cumhaz = -ln(km)
. line cumhaz residual residual, sort name("Weibull, replace") ///
> legend(off) t1(Weibull) xtitle("")
. drop cumhaz residual k
```

11. Calculate standardized residuals for the weibull model (post-estimation)

```
. stset futime,fail(fustat)id(id)
. xi:streg i.rx i.residds age  i.ecogps,dist(w)
```

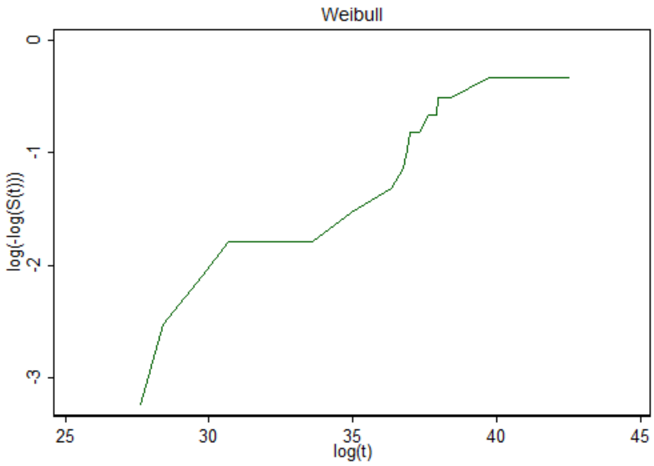




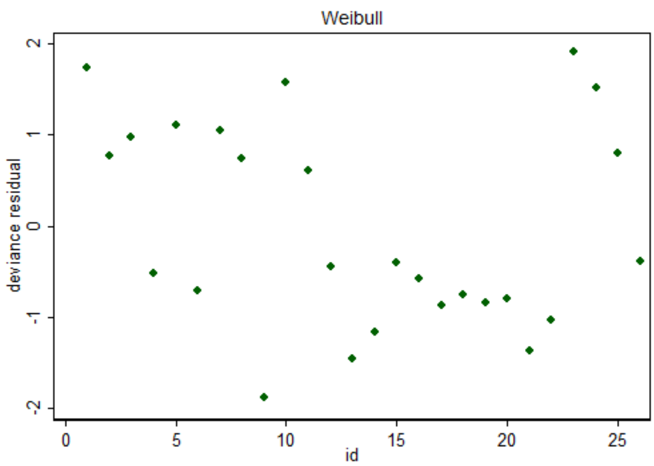

```
. predict pred, xb
. gen sigma= 0.5195874
. gen res_stand=(log(_t)-pred)/sigma
. replace res_stand=exp(res_stand)
. stset res_stand, failure(fustat)id(id)
. sts gen km = s
. gen gen tras_km=log(-log(km))
. gen log_time=log(_t)
. line tras_km log_time ,sort legend(off) t1(Weibull) xtitle("log(t)") ///
> ytitle("log(-log(S(t)))")
. drop km tras_km res_stand sigma pred log_time
```

12. Calculate deviance residuals for the weibull model (post-estimation)

```
. stset futime,fail(fustat)id(id)
. xi:streg i.rx i.residds age  i.ecogps,dist(w)
. predict deviance, deviance
```




```
. graph twoway scatter deviance id, legend(off) t1(Weibull) xtitle("id")
. drop deviance
```

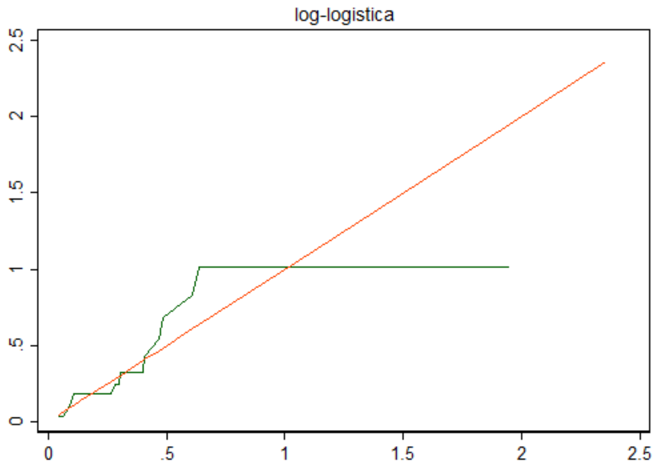


13. After fitting the log-logistic model, calculate Cox-Snell residuals (post-estimation)

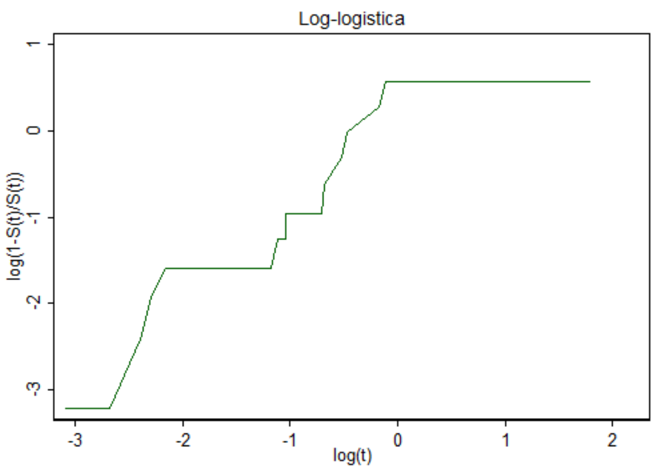
```
. stset futime,fail(fustat)id(id)
. xi:streg i.rx i.residds age  i.ecogps,dist(loglog)
. predict residual, csnell
. stset residual, failure(fustat)id(id)
. sts gen km = s
. line cumhaz residual residual, sort name("loglogistica, replace") ///
. legend(off) t1(log-logistica) xtitle("")
. drop cumhaz residual km
```

14. Calculate standardized residuals for the log-logistic model (post-estimation)

```
. stset futime,fail(fustat)id(id)
```



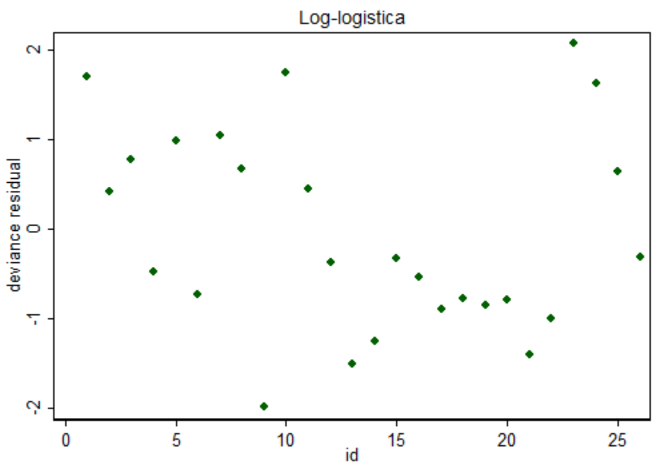

```
. xi:streg i.rx i.residds age  i.ecogps,dist(loglog)
. predict pred, xb
. gen sigma=.3727281
. gen res_stand=(log(_t)-pred)/sigma
. replace res_stand=exp(res_stand)
. stset res_stand, failure(fustat)id(id)
. sts gen km = s
. gen tras_km=log((1-km)/km)
. gen log_time=log(_t)
. line tras_km log_time,sort legend(off) t1(Log-logistica) xtitle("log(t)") ///
> ytitle("log(1-S(t)/S(t))")
. drop km tras_km res_stand sigma pred log_time
```



15. Calculate deviance residuals for the log-logistic model (post-estimation)

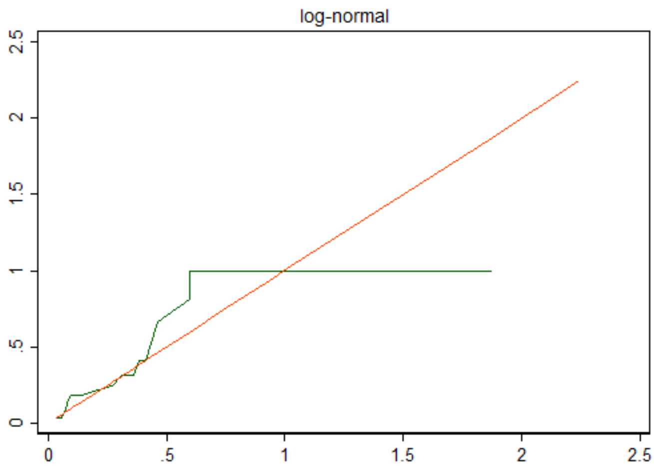
```
. stset futime,fail(fustat)id(id)
. xi:streg i.rx i.residds age  i.ecogps,dist(loglog)
```

```
. predict deviance, deviance
. graph twoway scatter deviance id, legend(off) t1(Log-logistica) xtitle("id")
. drop deviance
```



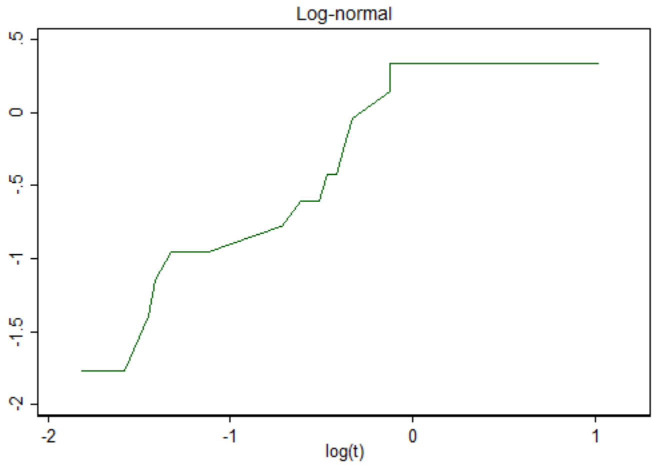
16. After fitting the log-normal model, calculate Cox-Snell residuals (post-estimation)

```
.stset futime,fail(fustat)id(id)
.xi:streg i.rx i.residds age  i.ecogps,dist(ln)
.predict residual, csnell
.stset residual, fail(fustat)id(id)
.sts gen km = s
.generate cumhaz = -ln(km)
.line cumhaz residual residual, sort name("lognormale, replace") ///
.legend(off) t1(log-normal) xtitle("")
.drop cumhaz residual km
```



17. Calculate standardized residuals for the log-normal model (post-estimation)

```
.stset futime,fail(fustat)id(id)
.xi:streg i.rx i.residds age  i.ecogps,dist(ln)
.predict pred, xb
.gen sigma=.6319339
.gen res_stand=(log(_t)-pred)/sigma
.replace res_stand=exp(res_stand)
.stset res_stand, fail(fustat)id(id)
.sts gen km = s
.gen tras_km=invnormal(1-km)
.gen log_time=log(_t)
.line tras_km log_time,sort legend(off) t1(Log-normal) xtitle("log(t)") ///
.ytitle("")
.drop km tras_km res_stand sigma pred log_time
```

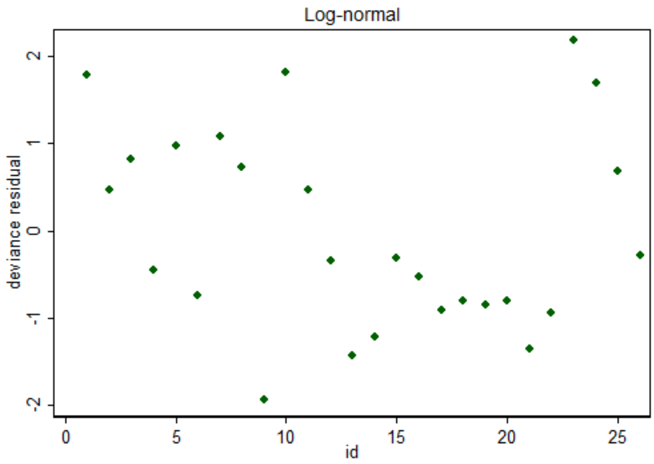


18. Calculate deviance residuals for the log-normal model (post-estimation)

```
stset futime,fail(fustat)id(id)
xi:streg i.rx i.residds age  i.ecogps,dist(ln)
predict deviance, deviance
graph twoway scatter deviance id, legend(off) t1(Log-normal) xtitle("id")
drop deviance
```

19. Compare the models by AIC (Akaike Criterion Information=2k-2ln(L))

```
stset futime,fail(fustat)id(id)
xi:streg i.rx i.residds age  i.ecogps,dist(e)
estat ic
-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
           . |     26   -29.84554   -22.45765      5      54.9153    61.20579
-----------------------------------------------------------------------------
xi:streg i.rx i.residds age  i.ecogps,dist(w)
estat ic
```




```
--------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+------------------------------------------------------------
           . |     26   -29.76724   -19.66043      6     51.32085    58.86943
--------------------------------------------------------------------------
```

```
xi:streg i.rx i.residds age  i.ecogps,dist(loglog)
estat ic
```

```
--------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+------------------------------------------------------------
           . |     26   -29.16804   -18.63665      6      49.2733    56.82188
--------------------------------------------------------------------------
```

```
xi:streg i.rx i.residds age  i.ecogps,dist(ln)
estat ic
```

```
--------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+------------------------------------------------------------
           . |     26   -28.93508   -18.22165      6      48.4433    55.99188
--------------------------------------------------------------------------
```

20. After selecting the best model fitting data, interpret the estimates obtained by that model

```
xi:streg i.rx i.residds age  i.ecogps,dist(ln)
```

```
Log likelihood  =   -18.22165                     Prob > chi2     =    0.0003
------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _Irx_1 |  -.6058923   .3026846    -2.00   0.045    -1.199143   -.0126413
  _Iresidds_2 |  -.5331212   .3493213    -1.53   0.127    -1.217778     .151536
          age |  -.0676027   .0162841    -4.15   0.000    -.0995189   -.0356865
   _Iecogps_2 |  -.0205032   .3203281    -0.06   0.949    -.6483347    .6073282
        _cons |   10.99911   .9817151    11.20   0.000     9.074987    12.92324
-------------+----------------------------------------------------------------
```

```
     /ln_sig |  -.4589705    .220193    -2.08   0.037    -.8905408   -.0274002
-------------+----------------------------------------------------------------
       sigma |   .6319339   .1391474                      .4104337    .9729718
------------------------------------------------------------------------------

xi:streg i.rx i.residds age  i.ecogps,dist(loglog)

Log likelihood  =    -18.63665                     Prob > chi2     =    0.0003

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      _Irx_1 |  -.5956984   .3039036    -1.96   0.050    -1.191338   -.0000584
 _Iresidds_2 |  -.5468831   .3530328    -1.55   0.121    -1.238815    .1450484
         age |  -.0690454   .0173885    -3.97   0.000    -.1031262   -.0349646
  _Iecogps_2 |  -.0345666   .3146566    -0.11   0.913    -.6512821    .5821489
       _cons |    11.0713   1.033283    10.71   0.000       9.0461     13.0965
-------------+----------------------------------------------------------------
     /ln_gam |  -.9869061   .2412497    -4.09   0.000    -1.459747   -.5140654
-------------+----------------------------------------------------------------
       gamma |   .3727281   .0899205                      .2322951    .5980593
------------------------------------------------------------------------------
```