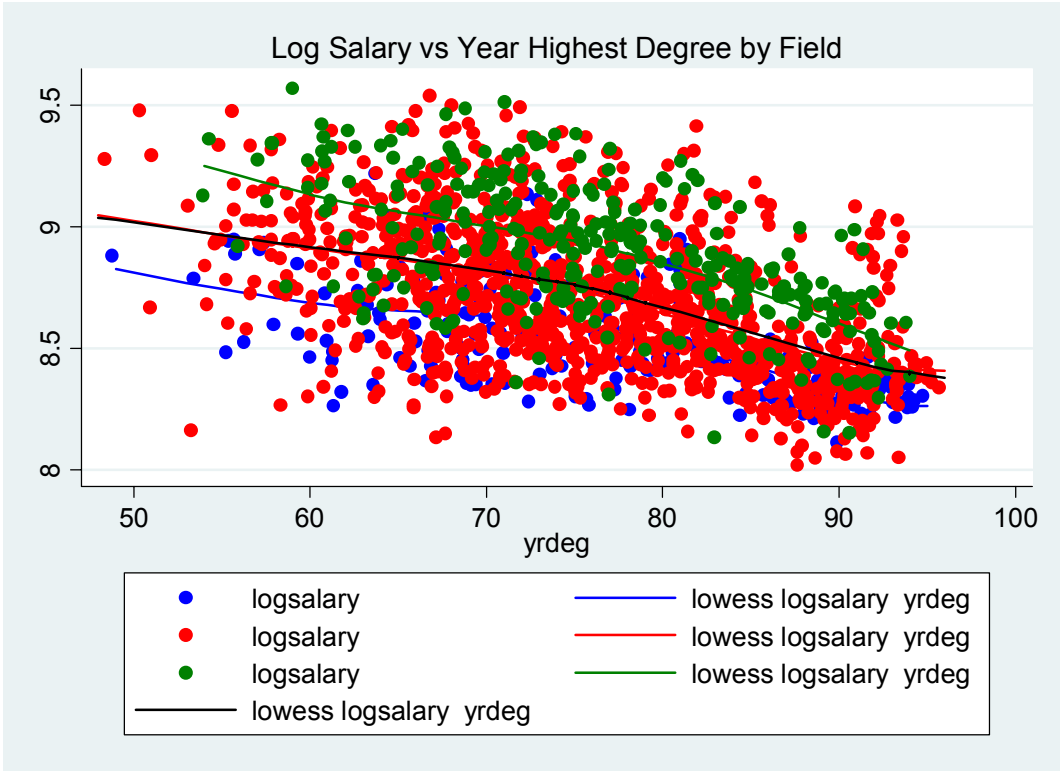
```
#### Biost 517: Applied Biostatistics I
#### Emerson, Fall 2005

#### Annotated Stata Log File: Homework #4
#### November 4, 2005

#### In this file I give the Stata commands I used to produce
#### the key to Homework #4. In order to properly format
#### a table useful to casual readers, I cut and pasted some
#### of the output into Excel.

#### Comments edited into the log file produced by Stata are
#### on the lines that start with the four '#' signs and are
#### printed in italics.

#### The Stata commands are put in bold face.

#### Stata output is displayed in regular typeface in blue.

-------------------------------------------------------------------------------
#### Open log file to save commands and results
. log using hw4Stata.log
-------------------------------------------------------------------------------
      log:  C:\sse\teach\b517\hw4Stata.log
  log type:  text
      log:  C:\My Documents\teach\b517\f05\hw4Stata.log
  log type:  text
 opened on:   2 Nov 2005, 13:38:57

#### The salary dataset is very large: I need to increase the memory available for the
#### data prior to reading in the dataset.
. set memory 24m
(24576k)

#### I had previously stored the salary dataset as a Stata data file. The commands used
#### to generate this dataset are in initsalary.doc.
. use ..\..\datasets\salary

#### Creating the variable containing the log transformed data
. gen logsalary= log(salary)

##################
####
#### Problem 1
####
##################

#### Create scatterplot and lowess curves stratified by field

. twoway (scatter logsalary yrdeg if year==95&field==1, col("blue") jitter(2)) (l
> owess logsalary yrdeg if year==95&field==1, col("blue")) (scatter logsalary yrd
> eg if year==95&field==2, col("red") jitter(2)) (lowess logsalary yrdeg if year=
> =95&field==2, col("red")) (scatter logsalary yrdeg if year==95&field==3, col("g
> reen") jitter(2)) (lowess logsalary yrdeg if year==95&field==3, col("green")) (
> lowess logsalary yrdeg if year==95, col("black")), t1("Log Salary vs Year Highe
> st Degree by Field")
```




```
#### Correlation between logsalary and yrdeg (this will be used in problems 1-3)
. cor logsalary yrdeg if year==95
(obs=1597)

             | logsal~y    yrdeg
-------------+------------------
   logsalary |   1.0000
       yrdeg |  -0.5101   1.0000


#### Correlation between logsalary and yrdeg within strata defined by field
. bysort field: cor logsalary yrdeg if year==95

-------------------------------------------------------------------------------
-> field = Arts
(obs=220)

             | logsal~y    yrdeg
-------------+------------------
   logsalary |   1.0000
       yrdeg |  -0.5826   1.0000


-------------------------------------------------------------------------------
-> field = Other
(obs=1067)

             | logsal~y    yrdeg
-------------+------------------
   logsalary |   1.0000
```

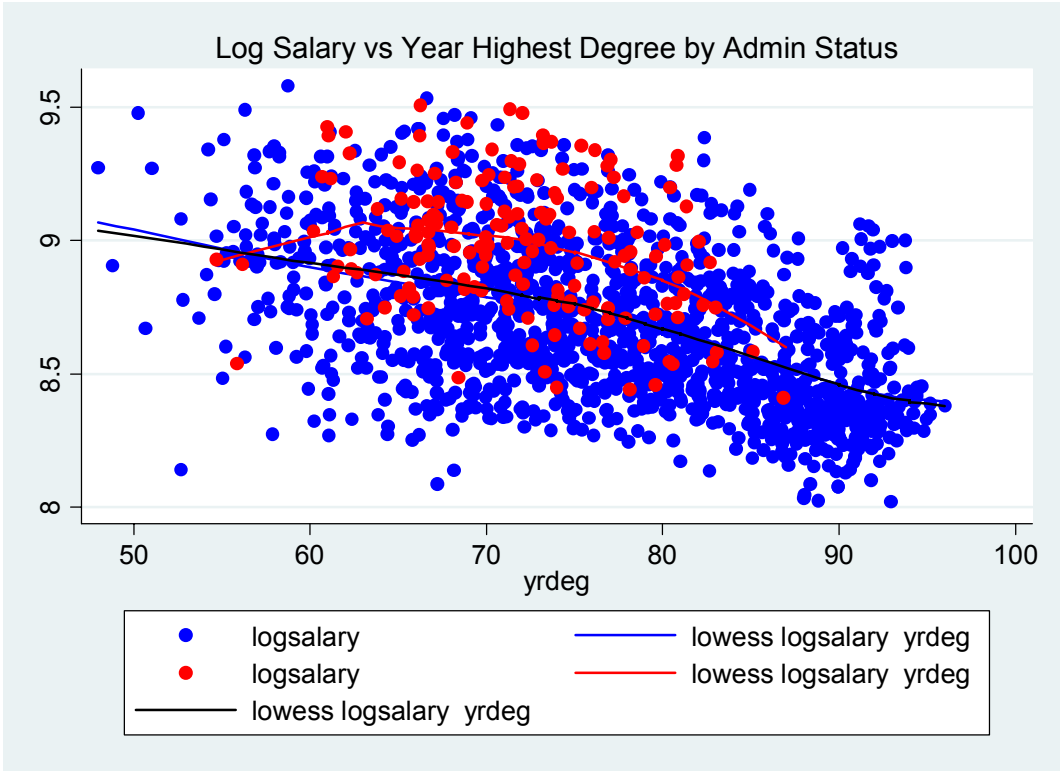
```
       yrdeg |  -0.5290   1.0000

-------------------------------------------------------------------------------
-> field = Prof
(obs=310)

             | logsal~y    yrdeg
-------------+------------------
   logsalary |   1.0000
       yrdeg |  -0.6035   1.0000


#### In order to be able to interpret the behavior of correlation overall and within
#### strata, I find it convenient to look at the standard deviation of yrdeg, the
#### least squares slope estimate of logsalary regressed on yrdeg, and the within
#### group (with groups defined by yrdeg) standard deviation of logsalary.
#### I look at these three statistics both overall and within strata defined by
#### field (for problem 1), admin (for problem 2), and rank (for problem 3).

#### Descriptive statistics of yrdeg (I am mainly interested in the sd)
. tabstat yrdeg if year==95, stat(n mean sd min p25 p50 p75 max) col(stat) format

    variable |    N  mean    sd       min       p25       p50       p75       max
-------------+--------------------------------------------------------------------
       yrdeg | 1597  76.1  9.86        48        69        76        84        96
-------------------------------------------------------------------------------

#### Descriptive statistics of yrdeg within strata defined by field
. bysort field: tabstat yrdeg if year==95, stat(n mean sd min p25 p50 p75 max)
col(stat) format

-------------------------------------------------------------------------------
-> field = Arts

    variable |    N  mean    sd       min       p25       p50       p75       max
-------------+--------------------------------------------------------------------
       yrdeg |  220  76.1  10.3        49        69        76      84.5        95
-------------------------------------------------------------------------------

-------------------------------------------------------------------------------
-> field = Other

    variable |    N   mean    sd       min       p25       p50       p75       max
-------------+--------------------------------------------------------------------
       yrdeg | 1067   76.0  9.90        48        68        75        84        96
-------------------------------------------------------------------------------

-------------------------------------------------------------------------------
-> field = Prof

    variable |    N  mean    sd       min       p25       p50       p75       max
-------------+--------------------------------------------------------------------
       yrdeg |  310  76.6  9.40        54        70        76        84        94
-------------------------------------------------------------------------------


#### Regression of logsalary on yrdeg: I am only interested in the slope (in the row marked
```

```
#### "yrdeg" and column marked "Coef.") and the estimated standard deviation of logsalary
#### within groups defined by yrdeg (labeled "Root MSE" in the output)
. regress logsalary yrdeg if year==95

      Source |       SS       df       MS              Number of obs =    1597
-------------+------------------------------           F(  1,  1595) =  560.86
       Model |  38.1176109     1  38.1176109           Prob > F      =  0.0000
    Residual |  108.399806  1595  .067962261           R-squared     =  0.2602
-------------+------------------------------           Adj R-squared =  0.2597
       Total |  146.517416  1596  .091802893           Root MSE      =   .2607

------------------------------------------------------------------------------
   logsalary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       yrdeg |  -.0156777    .000662   -23.68   0.000    -.0169761   -.0143792
       _cons |    9.90838   .0507946   195.07   0.000     9.808749    10.00801
------------------------------------------------------------------------------

#### Regression of logsalary on yrdeg within strata defined by field
. bysort field: regress logsalary yrdeg if year==95

-------------------------------------------------------------------------------
-> field = Arts

      Source |       SS       df       MS              Number of obs =     220
-------------+------------------------------           F(  1,   218) =  112.01
       Model |   3.7689177     1   3.7689177           Prob > F      =  0.0000
    Residual |  7.33528964   218  .033648118           R-squared     =  0.3394
-------------+------------------------------           Adj R-squared =  0.3364
       Total |  11.1042073   219  .050704143           Root MSE      =  .18343

------------------------------------------------------------------------------
   logsalary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       yrdeg |  -.0127713   .0012067   -10.58   0.000    -.0151496    -.010393
       _cons |   9.517413   .0926985   102.67   0.000     9.334713    9.700113
------------------------------------------------------------------------------

-------------------------------------------------------------------------------
-> field = Other

      Source |       SS       df       MS              Number of obs =    1067
-------------+------------------------------           F(  1,  1065) =  413.77
       Model |  26.8460588     1  26.8460588           Prob > F      =  0.0000
    Residual |  69.0989521  1065  .064881645           R-squared     =  0.2798
-------------+------------------------------           Adj R-squared =  0.2791
       Total |  95.9450109  1066  .090004701           Root MSE      =  .25472

------------------------------------------------------------------------------
   logsalary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       yrdeg |  -.0160218   .0007876   -20.34   0.000    -.0175673   -.0144763
       _cons |   9.917613   .0603296   164.39   0.000     9.799235    10.03599
------------------------------------------------------------------------------

-------------------------------------------------------------------------------
-> field = Prof
```

```
      Source |       SS       df       MS              Number of obs =     310
-------------+------------------------------           F(  1,   308) =  176.45
       Model |  8.65328039     1  8.65328039            Prob > F      =  0.0000
    Residual |  15.1047021   308  .049041241           R-squared     =  0.3642
-------------+------------------------------           Adj R-squared =  0.3622
       Total |  23.7579825   309  .076886675           Root MSE      =  .22145

------------------------------------------------------------------------------
   logsalary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       yrdeg |  -.0178022   .0013402   -13.28   0.000    -.0204393   -.0151651
       _cons |   10.24964   .1033698    99.16   0.000     10.04624    10.45304
------------------------------------------------------------------------------
```

```
##################
####
#### Problem 2
####
##################

#### Create scatterplot and lowess curves stratified by administrative duties
```

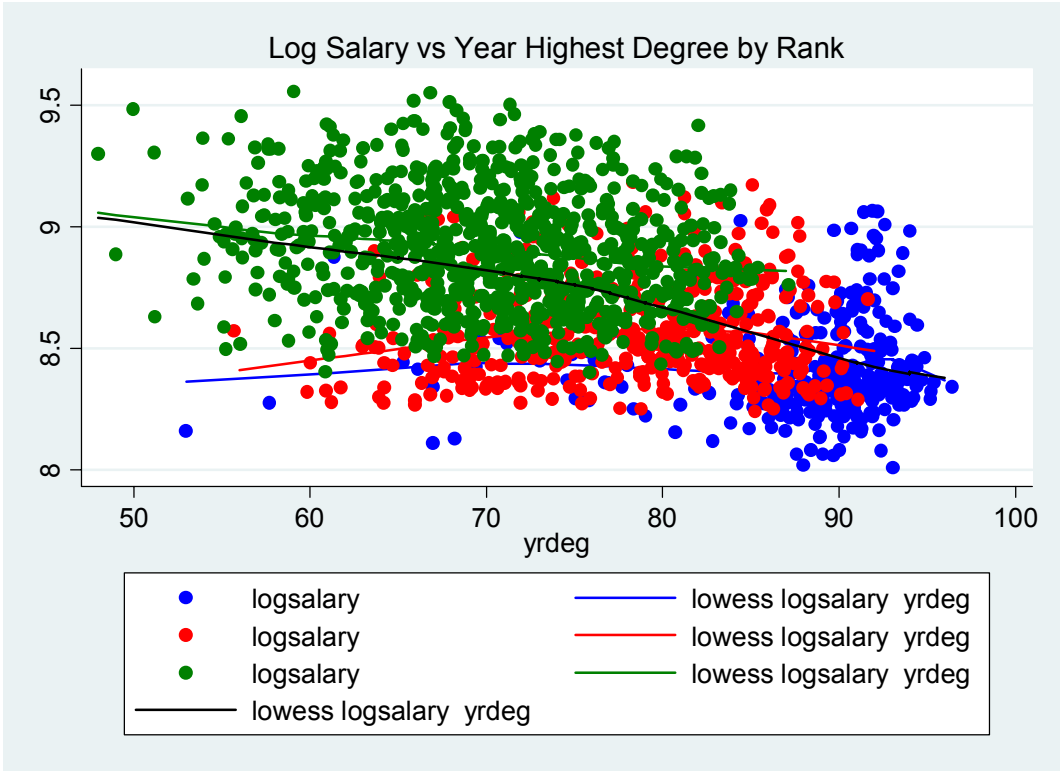
```
. twoway (scatter logsalary yrdeg if year==95&admin==0, col("blue") jitter(2)) (l
> owess logsalary yrdeg if year==95&admin==0, col("blue")) (scatter logsalary yrd
> eg if year==95&admin==1, col("red") jitter(2)) (lowess logsalary yrdeg if year=
> =95&admin==1, col("red"))  (lowess logsalary yrdeg if year==95, col("black")),
> t1("Log Salary vs Year Highest Degree by Admin Status")
```



```
#### Correlation between logsalary and yrdeg within strata defined by admin
. bysort admin: cor logsalary yrdeg if year==95

-------------------------------------------------------------------------------
-> admin = Nonadmin
(obs=1428)

             | logsal~y    yrdeg
-------------+------------------
   logsalary |   1.0000
       yrdeg |  -0.5061   1.0000


-------------------------------------------------------------------------------
-> admin = Admin
(obs=169)

             | logsal~y    yrdeg
-------------+------------------
   logsalary |   1.0000
       yrdeg |  -0.2865   1.0000


#### Descriptive statistics of yrdeg within strata defined by admin
. bysort admin: tabstat yrdeg if year==95, stat(n mean sd min p25 p50 p75 max)
col(stat) format

-------------------------------------------------------------------------------
-> admin = Nonadmin

    variable |    N  mean    sd       min       p25       p50       p75       max
-------------+--------------------------------------------------------------------
       yrdeg | 1428  76.6  10.1        48        69        77        85        96
-------------------------------------------------------------------------------

-------------------------------------------------------------------------------
-> admin = Admin

    variable |    N  mean    sd       min       p25       p50       p75       max
-------------+--------------------------------------------------------------------
       yrdeg |  169  71.7  6.30        55        67        72        76        87
-------------------------------------------------------------------------------

#### Regression of logsalary on yrdeg within strata defined by admin
. bysort admin: regress logsalary yrdeg if year==95

-------------------------------------------------------------------------------
-> admin = Nonadmin

      Source |       SS       df       MS              Number of obs =    1428
-------------+------------------------------           F(  1,  1426) =  491.00
       Model |  31.6887618     1  31.6887618           Prob > F      =  0.0000
    Residual |  92.0325576  1426   .06453896           R-squared     =  0.2561
-------------+------------------------------           Adj R-squared =  0.2556
       Total |  123.721319  1427  .086700294           Root MSE      =  .25405
```

```
------------------------------------------------------------------------------
   logsalary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       yrdeg |   -.014795   .0006677   -22.16   0.000    -.0161048   -.0134853
       _cons |   9.818217   .0515943   190.30   0.000     9.717009    9.919426
------------------------------------------------------------------------------

------------------------------------------------------------------------------
-> admin = Admin

      Source |       SS       df       MS              Number of obs =     169
-------------+------------------------------           F(  1,   167) =   14.93
       Model |  .828020751     1  .828020751           Prob > F      =  0.0002
    Residual |  9.26108304   167  .055455587           R-squared     =  0.0821
-------------+------------------------------           Adj R-squared =  0.0766
       Total |  10.0891038   168  .060054189           Root MSE      =  .23549

------------------------------------------------------------------------------
   logsalary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       yrdeg |  -.0111409   .0028832    -3.86   0.000     -.016833   -.0054487
       _cons |   9.773533   .2075274    47.10   0.000     9.363818    10.18325
------------------------------------------------------------------------------
```

```
##################
####
#### Problem 3
####
##################

#### Create scatterplot and lowess curves stratified by rank

. twoway (scatter logsalary yrdeg if year==95&rank==1, col("blue") jitter(2)) (lo
> wess logsalary yrdeg if year==95&rank==1, col("blue")) (scatter logsalary yrdeg
>  if year==95&rank==2, col("red") jitter(2)) (lowess logsalary yrdeg if year==95
> &rank==2, col("red")) (scatter logsalary yrdeg if year==95&rank==3, col("green"
> ) jitter(2)) (lowess logsalary yrdeg if year==95&rank==3, col("green")) (lowess
>  logsalary yrdeg if year==95, col("black")), t1("Log Salary vs Year Highest Deg
> ree by Rank")
```




```
#### Correlation between logsalary and yrdeg within strata defined by rank
. bysort rank: cor logsalary yrdeg if year==95

-------------------------------------------------------------------------------
-> rank = Assist
(obs=315)

             | logsal~y    yrdeg
-------------+------------------
   logsalary |   1.0000
       yrdeg |   0.0457   1.0000


-------------------------------------------------------------------------------
-> rank = Assoc
(obs=437)

             | logsal~y    yrdeg
-------------+------------------
   logsalary |   1.0000
       yrdeg |   0.0520   1.0000


-------------------------------------------------------------------------------
-> rank = Full
(obs=845)

             | logsal~y    yrdeg
-------------+------------------
   logsalary |   1.0000
```

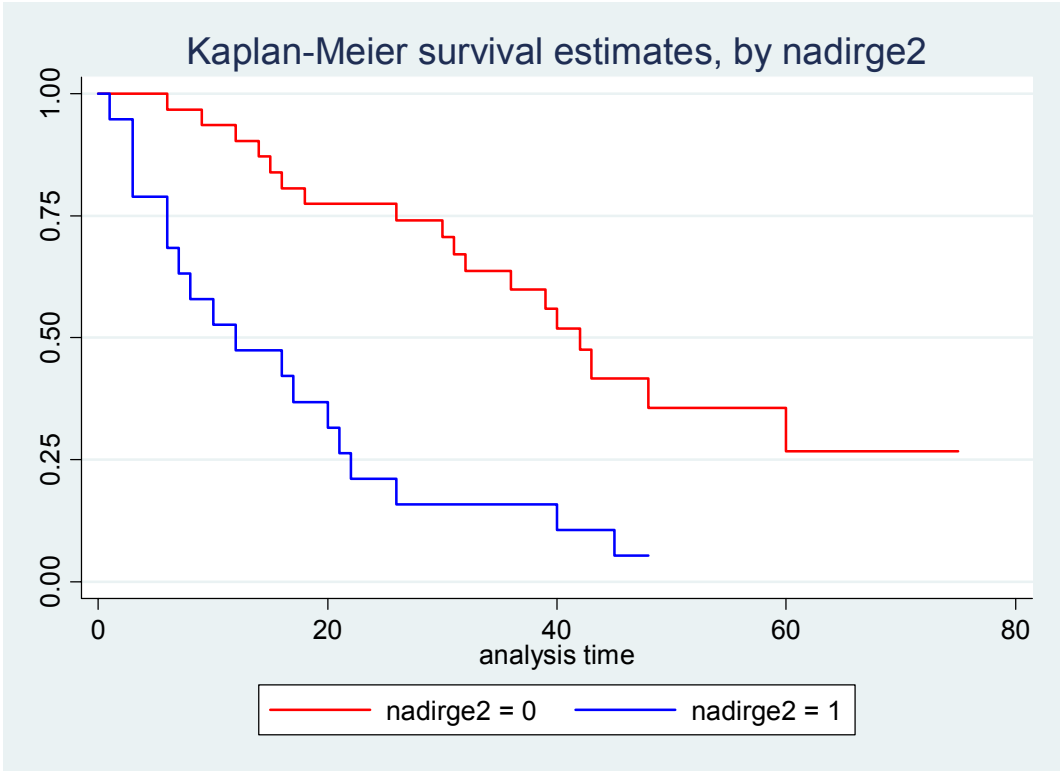
```
       yrdeg |  -0.1978   1.0000

-------------------------------------------------------------------------------
-> rank = .
no observations

#### Descriptive statistics of yrdeg within strata defined by rank
. bysort rank: tabstat yrdeg if year==95, stat(n mean sd min p25 p50 p75 max)
col(stat) format

-------------------------------------------------------------------------------
-> rank = Assist

    variable |    N  mean    sd       min       p25       p50       p75       max
-------------+--------------------------------------------------------------------
       yrdeg |  315  88.0  6.37        53        87        90        92        96
-------------------------------------------------------------------------------

-------------------------------------------------------------------------------
-> rank = Assoc

    variable |    N  mean    sd       min       p25       p50       p75       max
-------------+--------------------------------------------------------------------
       yrdeg |  437  78.6  7.20        56        73        80        85        92
-------------------------------------------------------------------------------

-------------------------------------------------------------------------------
-> rank = Full

    variable |    N  mean    sd       min       p25       p50       p75       max
-------------+--------------------------------------------------------------------
       yrdeg |  845  70.3  7.23        48        66        71        75        87
-------------------------------------------------------------------------------

-------------------------------------------------------------------------------
-> rank = .
no observations

#### Regression of logsalary on yrdeg within strata defined by rank
. bysort rank: regress logsalary yrdeg if year==95

-------------------------------------------------------------------------------
-> rank = Assist

      Source |       SS       df       MS              Number of obs =     315
-------------+------------------------------           F(  1,   313) =    0.65
       Model |  .027404289     1  .027404289           Prob > F      =  0.4194
    Residual |  13.1183403   313   .04191163           R-squared     =  0.0021
-------------+------------------------------           Adj R-squared = -0.0011
       Total |  13.1457445   314  .041865428           Root MSE      =  .20472

------------------------------------------------------------------------------
   logsalary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       yrdeg |   .0014657   .0018126     0.81   0.419    -.0021008    .0050322
       _cons |   8.293092   .1599515    51.85   0.000     7.978375    8.607808
```

```
-------------------------------------------------------------------------------

-------------------------------------------------------------------------------
-> rank = Assoc

      Source |       SS       df       MS              Number of obs =     437
-------------+------------------------------           F(  1,   435) =    1.18
       Model |  .045216708     1  .045216708           Prob > F      =  0.2781
    Residual |  16.6785717   435  .038341544           R-squared     =  0.0027
-------------+------------------------------           Adj R-squared =  0.0004
       Total |  16.7237884   436  .038357313           Root MSE      =  .19581

------------------------------------------------------------------------------
   logsalary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       yrdeg |   .0014141   .0013021     1.09   0.278    -.0011452    .0039734
       _cons |    8.45058   .1028292    82.18   0.000     8.248476    8.652684
------------------------------------------------------------------------------

-------------------------------------------------------------------------------
-> rank = Full

      Source |       SS       df       MS              Number of obs =     845
-------------+------------------------------           F(  1,   843) =   34.34
       Model |  1.92260796     1  1.92260796           Prob > F      =  0.0000
    Residual |  47.2019378   843  .055992809           R-squared     =  0.0391
-------------+------------------------------           Adj R-squared =  0.0380
       Total |  49.1245457   844  .058204438           Root MSE      =  .23663

------------------------------------------------------------------------------
   logsalary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       yrdeg |  -.0066023   .0011267    -5.86   0.000    -.0088138   -.0043908
       _cons |   9.368547   .0796648   117.60   0.000     9.212182    9.524911
------------------------------------------------------------------------------

-------------------------------------------------------------------------------
-> rank = .
no observations


#### Clearing this dataset, and then reading in the PSA dataset
. clear

. infile ptid nadirpsa pretxpsa ps bss grade age obstime str8 inrem using  ..\..\
> datasets\2005.09.26\psa.txt
'ptid' cannot be read as a number for ptid[1]
'nadirpsa' cannot be read as a number for nadirpsa[1]
'pretxpsa' cannot be read as a number for pretxpsa[1]
'ps' cannot be read as a number for ps[1]
'bss' cannot be read as a number for bss[1]
'grade' cannot be read as a number for grade[1]
'age' cannot be read as a number for age[1]
'obstime' cannot be read as a number for obstime[1]
'NA' cannot be read as a number for pretxpsa[8]
'NA' cannot be read as a number for pretxpsa[15]
'NA' cannot be read as a number for pretxpsa[18]
```

```
'NA' cannot be read as a number for grade[22]
'NA' cannot be read as a number for grade[24]
'NA' cannot be read as a number for grade[26]
'NA' cannot be read as a number for grade[27]
'NA' cannot be read as a number for grade[30]
'NA' cannot be read as a number for grade[33]
'NA' cannot be read as a number for pretxpsa[35]
'NA' cannot be read as a number for grade[35]
'NA' cannot be read as a number for ps[37]
'NA' cannot be read as a number for bss[37]
'NA' cannot be read as a number for pretxpsa[43]
'NA' cannot be read as a number for ps[43]
'NA' cannot be read as a number for bss[43]
'NA' cannot be read as a number for grade[43]
'NA' cannot be read as a number for pretxpsa[46]
'NA' cannot be read as a number for pretxpsa[51]
'NA' cannot be read as a number for grade[51]
(51 observations read)

#### Verifying that I can safely drop the first case.
. list in 1

     +-------------------------------------------------------------------------+
     | ptid   nadirpsa   pretxpsa   ps   bss   grade   age   obstime   inrem |
     |-------------------------------------------------------------------------|
  1. |    .          .          .    .     .       .     .         .   inrem |
     +-------------------------------------------------------------------------+

. drop in 1
(1 observation deleted)

##################
####
#### Problem 4
####
##################

#### Create variables indicating relapse, relapse within 24 months, and dichotomization
#### of nadirpsa at 2 ng/ml and 4 ng/ml
. gen relapse= 0

. replace relapse=1 if inrem=="no"
(36 real changes made)

. gen relaps24= 0

. replace relaps24=1 if inrem=="no" & obstime <=24
(22 real changes made)

. gen nadirge2= nadirpsa

. recode nadirge2 min/2=0 2/max=1
(nadirge2: 50 changes made)

. gen nadirge4= nadirpsa

. recode nadirge4 min/4=0 4/max=1
(nadirge4: 50 changes made)
```

```
#### "Setting" the response variables for Kaplan-Meier estimation
. stset obstime relapse

     failure event:  relapse != 0 & relapse < .
obs. time interval:  (0, obstime]
 exit on or before:  failure

--------------------------------------------------------------------------------
       50  total obs.
        0  exclusions
--------------------------------------------------------------------------------
       50  obs. remaining, representing
       36  failures in single record/single failure data
     1423  total analysis time at risk, at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =        75

#### Listing the KM estimates to get descriptive statistics: I will give estimated
#### quartiles, median, and survival probabilities by year.
####
#### The pth percentile will be the time at which the Survivor Function first goes
#### below p (If the Survivor Function is equal to p exactly, I will take the midpoint
#### between the time at which the Survivor Function equals p and the next time the
#### Survivor Function changes).
####       Example: To get the estimated time at which the probability of remaining
####                in remission is 0.75, I choose 12 months, because that is the
####                first time the Survivor Function is estimated to be less than 0.75
####       Example: To get the estimated time at which the probability of remaining
####                in remission is 0.70, I choose 15.5 months, because the estimated
####                Survivor Function is 0.70 from 15 months to 16 months (when it changes
####                to 0.66).
####
#### To get the 12, 24, 36, etc. month survival probability, I will look at the largest
#### time equal to or less than the desired threshold.
####       Example: The estimated probability of remaining in remission for more than 12
####                months is 0.74 (an estimate given explicitly in the listing)
####       Example: The estimated probability of remaining in remission for more that 23
####                months is 0.56 (using the estimate given for 22 months, because
####                the Survivor Function estimate does not change between 22 and 24
####                months)

. sts list

         failure _d:  relapse
   analysis time _t:  obstime

          Beg.          Net            Survivor      Std.
  Time    Total   Fail   Lost           Function     Error     [95% Conf. Int.]
-------------------------------------------------------------------------------
     1       50      1      0             0.9800    0.0198     0.8664    0.9972
     3       49      3      0             0.9200    0.0384     0.8007    0.9692
     6       46      3      0             0.8600    0.0491     0.7286    0.9307
     7       43      1      0             0.8400    0.0518     0.7054    0.9166
     8       42      1      0             0.8200    0.0543     0.6826    0.9020
     9       41      1      0             0.8000    0.0566     0.6602    0.8870
    10       40      1      0             0.7800    0.0586     0.6381    0.8716
    12       39      2      0             0.7400    0.0620     0.5947    0.8399
    14       37      1      0             0.7200    0.0635     0.5735    0.8236
```

```
    15        36       1       0            0.7000    0.0648     0.5525    0.8070
    16        35       2       0            0.6600    0.0670     0.5114    0.7730
    17        33       1       0            0.6400    0.0679     0.4911    0.7557
    18        32       1       0            0.6200    0.0686     0.4711    0.7381
    20        31       1       0            0.6000    0.0693     0.4513    0.7204
    21        30       1       0            0.5800    0.0698     0.4318    0.7024
    22        29       1       0            0.5600    0.0702     0.4124    0.6842
    24        28       0       1            0.5600    0.0702     0.4124    0.6842
    26        27       2       0            0.5185    0.0709     0.3725    0.6461
    30        25       1       1            0.4978    0.0710     0.3529    0.6267
    31        23       1       0            0.4761    0.0711     0.3325    0.6065
    32        22       1       0            0.4545    0.0711     0.3124    0.5860
    35        21       0       1            0.4545    0.0711     0.3124    0.5860
    36        20       1       1            0.4318    0.0711     0.2913    0.5645
    39        18       1       0            0.4078    0.0711     0.2691    0.5418
    40        17       2       1            0.3598    0.0703     0.2260    0.4953
    42        14       1       3            0.3341    0.0699     0.2033    0.4703
    43        10       1       0            0.3007    0.0704     0.1719    0.4404
    45         9       1       0            0.2673    0.0701     0.1426    0.4091
    48         8       1       2            0.2339    0.0688     0.1152    0.3765
    58         5       0       1            0.2339    0.0688     0.1152    0.3765
    60         4       1       2            0.1754    0.0723     0.0627    0.3348
    75         1       0       1            0.1754    0.0723     0.0627    0.3348
------------------------------------------------------------------------------


#### Graph of Kaplan Meier curves stratified by nadirge2
. sts graph, by(nadirge2) col("red" "blue")

         failure _d:  relapse
   analysis time _t:  obstime
```



Listing of Kaplan Meier curves stratified by nadirge2 to get quartiles

```
. sts list, by(nadirge2)

         failure _d:  relapse
   analysis time _t:  obstime
```

Time	Beg. Total	Fail	Net Lost	Survivor Function	Std. Error	[95% Conf.	Int.]
nadirge2=0							
6	31	1	0	0.9677	0.0317	0.7923	0.9954
9	30	1	0	0.9355	0.0441	0.7659	0.9835
12	29	1	0	0.9032	0.0531	0.7293	0.9677
14	28	1	0	0.8710	0.0602	0.6919	0.9495
15	27	1	0	0.8387	0.0661	0.6550	0.9295
16	26	1	0	0.8065	0.0710	0.6191	0.9080
18	25	1	0	0.7742	0.0751	0.5840	0.8854
24	24	0	1	0.7742	0.0751	0.5840	0.8854
26	23	1	0	0.7405	0.0790	0.5474	0.8609
30	22	1	1	0.7069	0.0823	0.5119	0.8355
31	20	1	0	0.6715	0.0854	0.4748	0.8082
32	19	1	0	0.6362	0.0879	0.4390	0.7800
35	18	0	1	0.6362	0.0879	0.4390	0.7800
36	17	1	1	0.5988	0.0904	0.4013	0.7497
39	15	1	0	0.5588	0.0927	0.3614	0.7170
40	14	1	1	0.5189	0.0943	0.3233	0.6831
42	12	1	3	0.4757	0.0959	0.2824	0.6462
43	8	1	0	0.4162	0.1006	0.2220	0.6002
48	7	1	1	0.3568	0.1023	0.1689	0.5503
58	5	0	1	0.3568	0.1023	0.1689	0.5503

```
    60        4      1      2             0.2676    0.1089     0.0894    0.4868
    75        1      0      1             0.2676    0.1089     0.0894    0.4868
nadirge2=1
     1       19      1      0             0.9474    0.0512     0.6812    0.9924
     3       18      3      0             0.7895    0.0935     0.5319    0.9153
     6       15      2      0             0.6842    0.1066     0.4279    0.8439
     7       13      1      0             0.6316    0.1107     0.3790    0.8044
     8       12      1      0             0.5789    0.1133     0.3321    0.7626
    10       11      1      0             0.5263    0.1145     0.2872    0.7188
    12       10      1      0             0.4737    0.1145     0.2444    0.6728
    16        9      1      0             0.4211    0.1133     0.2037    0.6249
    17        8      1      0             0.3684    0.1107     0.1652    0.5748
    20        7      1      0             0.3158    0.1066     0.1291    0.5225
    21        6      1      0             0.2632    0.1010     0.0958    0.4677
    22        5      1      0             0.2105    0.0935     0.0656    0.4102
    26        4      1      0             0.1579    0.0837     0.0392    0.3494
    40        3      1      0             0.1053    0.0704     0.0178    0.2843
    45        2      1      0             0.0526    0.0512     0.0036    0.2143
    48        1      0      1             0.0526    0.0512     0.0036    0.2143
-------------------------------------------------------------------------------
```

Listing of Kaplan Meier curves at 12, 24, 36, and 48 mos stratified by nadirge2

```
. sts list, by(nadirge2) at(12 24 36 48)

         failure _d:  relapse
   analysis time _t:  obstime

              Beg.                    Survivor      Std.
    Time     Total     Fail            Function     Error     [95% Conf. Int.]
-------------------------------------------------------------------------------
nadirge2=0
      12        29        3             0.9032    0.0531     0.7293    0.9677
      24        24        4             0.7742    0.0751     0.5840    0.8854
      36        17        5             0.5988    0.0904     0.4013    0.7497
      48         7        5             0.3568    0.1023     0.1689    0.5503
nadirge2=1
      12        10       10             0.4737    0.1145     0.2444    0.6728
      24         5        5             0.2105    0.0935     0.0656    0.4102
      36         4        1             0.1579    0.0837     0.0392    0.3494
      48         1        2             0.0526    0.0512     0.0036    0.2143
-------------------------------------------------------------------------------
Note:  Survivor function is calculated over full data and evaluated at
       indicated times; it is not calculated from aggregates shown at left.
```

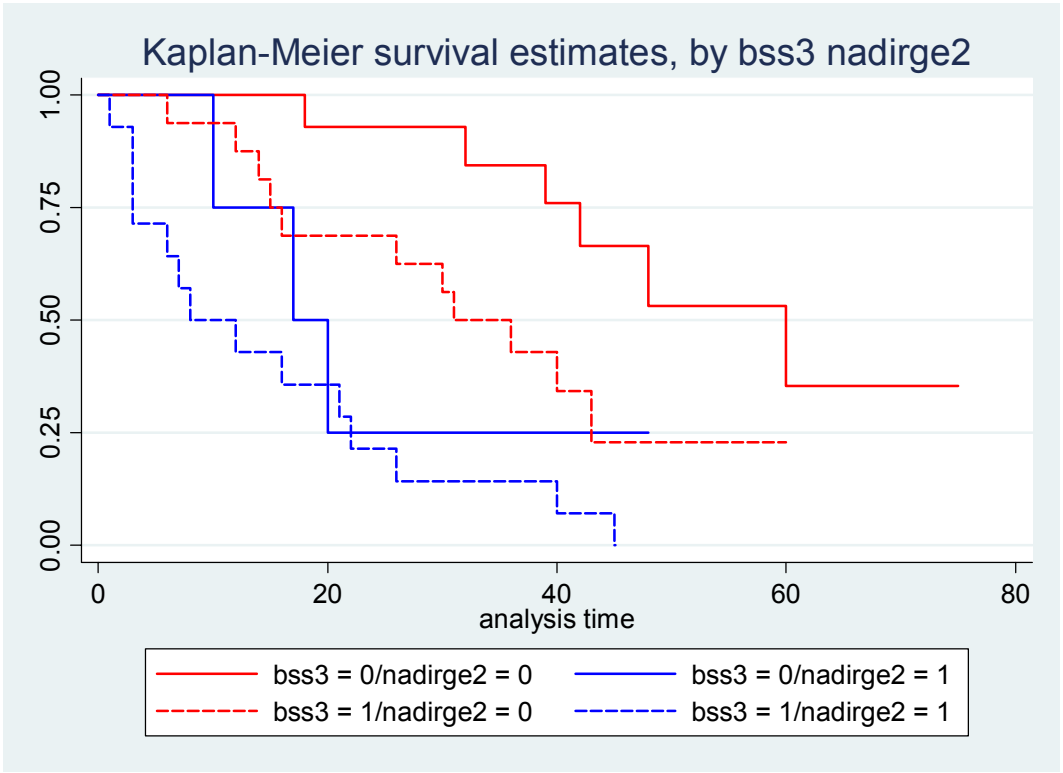
In order to consider the association between relapse and nadirpsa that is independent
of bone scan score, I will want to adjust for bss. Since bss is categorical, I
choose to just see whether there is an association between relapse and nadirpsa
within strata defined by bss. I examine the number of events in each bss stratum,
because it is not worth doing an analysis within a stratum that has too few
observed events

Examining the number of events within bss categories. There is only one observed
relapse in the bss==1 stratum, so I decide to consider the association between
relapse and nadirpsa within two strata defined by whether bss==3 or not.

```
. table bss relapse

----------------------
          |  relapse
```

```
      bss |     0      1
----------+-----------
        1 |     4      1
        2 |     5      8
        3 |     5     25
----------------------
```

```
#### Creating the dichotomized variables for bone scan score >= 3.
#### Note that I had to consider the missing data explicitly when I
#### created the variable this way ("recode" would have handled
#### it automatically).
. gen bss3=0

. replace bss3=1 if bss==3
(30 real changes made)

. replace bss3=. if bss==.
(2 real changes made, 2 to missing)


#### Graph of Kaplan Meier stratified by nadirge2 and bss3
. sts graph, by(bss3 nadirge2) col("red" "blue" "red" "blue") lp("l" "l" "-" "-")

         failure _d:  relapse
   analysis time _t:  obstime
```




```
#### Listing of Kaplan Meier at specific times stratified by nadirge2 and bss3
. sts list, by(bss3 nadirge2) at(12 24 36 48)

         failure _d:  relapse
```

```
   analysis time _t:  obstime

             Beg.                          Survivor      Std.
    Time     Total     Fail                Function     Error     [95% Conf. Int.]
-------------------------------------------------------------------------------
bss3=0 nadirge2=0
      12         0        0                1.0000         .          .          .
      24        13        1                0.9286    0.0688     0.5908     0.9896
      36        11        1                0.8442    0.1019     0.5041     0.9590
      48         5        3                0.5318    0.1627     0.1956     0.7832
bss3=0 nadirge2=1
      12         4        1                0.7500    0.2165     0.1279     0.9605
      24         2        2                0.2500    0.2165     0.0089     0.6653
      36         2        0                0.2500    0.2165     0.0089     0.6653
      48         1        0                0.2500    0.2165     0.0089     0.6653
bss3=1 nadirge2=0
      12        15        2                0.8750    0.0827     0.5860     0.9672
      24        12        3                0.6875    0.1159     0.4046     0.8563
      36         7        4                0.4286    0.1259     0.1879     0.6509
      48         2        2                0.2286    0.1258     0.0466     0.4914
bss3=1 nadirge2=1
      12         7        8                0.4286    0.1323     0.1773     0.6604
      24         4        3                0.2143    0.1097     0.0521     0.4479
      36         3        1                0.1429    0.0935     0.0232     0.3655
      48         1        2                     .         .          .          .
-------------------------------------------------------------------------------
Note:  Survivor function is calculated over full data and evaluated at
       indicated times; it is not calculated from aggregates shown at left.


##################
####
#### Problem 5
####
##################

#### Estimates of prevalence of relapse within 24 months
. ci relaps24, binomial

                                                      -- Binomial Exact --
    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
    relaps24 |         50         .44    .0701997        .2999072    .5874559

#### Estimates of 1-specificity (when relaps24==0) and sensitivity (when relaps24==1)
#### of a nadirpsa greater than or equal to 2.
. bysort relaps24: ci nadirge2, binomial

-------------------------------------------------------------------------------
-> relaps24 = 0

                                                      -- Binomial Exact --
    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
    nadirge2 |         28    .1428571      .06613        .0403356    .3266527

-------------------------------------------------------------------------------
-> relaps24 = 1
```

```
                                                        -- Binomial Exact --
    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
    nadirge2 |         22    .6818182    .0993026        .4512756    .8613535

#### Presuming that the observed data represents a reasonable cross-sectional sample,
#### we can estimate the positive predictive value and negative predictive value from
#### the sample.

#### Estimates of 1-neg pred value (when nadirge2==0) and pos pred value (when nadirge2==1)
#### of a nadirpsa greater than or equal to 2 predicting relapse within 24 months.
. bysort nadirge2: ci relaps24, binomial

-------------------------------------------------------------------------------
-> nadirge2 = 0

                                                        -- Binomial Exact --
    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
    relaps24 |         31    .2258065    .0750952        .0959422    .4109639

-------------------------------------------------------------------------------
-> nadirge2 = 1

                                                        -- Binomial Exact --
    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
    relaps24 |         19    .7894737    .0935288        .5443469    .9394755


#### Estimates of 1-specificity (when relaps24==0) and sensitivity (when relaps24==1)
#### of a nadirpsa greater than or equal to 4.
. bysort relaps24: ci nadirge4, binomial

-------------------------------------------------------------------------------
-> relaps24 = 0

                                                        -- Binomial Exact --
    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
    nadirge4 |         28    .0714286    .0486704        .0087705    .2350348

-------------------------------------------------------------------------------
-> relaps24 = 1

                                                        -- Binomial Exact --
    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
    nadirge4 |         22    .6818182    .0993026        .4512756    .8613535

#### Presuming that the observed data represents a reasonable cross-sectional sample,
#### we can estimate the positive predictive value and negative predictive value from
#### the sample.

#### Estimates of 1-neg pred value (when nadirge2==0) and pos pred value (when nadirge2==1)
#### of a nadirpsa greater than or equal to 2 predicting relapse within 24 months.
. bysort nadirge4: ci relaps24, binomial
```

```
-------------------------------------------------------------------------------
-> nadirge4 = 0

                                                    -- Binomial Exact --
    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
    relaps24 |         33    .2121212    .0711647        .0898042    .3890809

-------------------------------------------------------------------------------
-> nadirge4 = 1

                                                    -- Binomial Exact --
    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
    relaps24 |         17    .8823529    .0781425        .6355908    .9854207

. log close
      log:  C:\My Documents\teach\b517\f05\hw4Stata.log
  log type:  text
 closed on:   2 Nov 2005, 15:54:50
```