```
#### Biost 518: Applied Biostatistics II
#### Emerson, Winter 2008

#### Homework #3 Key
#### Annotated Stata Log File
#### February 9, 2008

#### The following output was used to generate the numbers that I wanted to present
#### in tables, as well as the plots I wanted to present as figures in the paper.
#### I note that Stata does not present its output in a form suitable for presentation.
#### Numbers need to be rounded to an interesting number of significant digits, and
#### the columns and rows need to facilitate comparison of relevant measures.
#### I used Excel to manipulate this output into the form I wanted, then copied the
#### resulting tables to the MS-Word document.

#### 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.

#### Data was read in and processed according to the leukemia.do file.

###############################################################################################
#### Problem 1: Analysis of induction of complete remission by treatment and sex
###############################################################################################

#### Problem 1a: Descriptive statistics

. tabulate male cr, row
| Key            |
|   frequency    |
| row percentage |
            |          cr
       male |         0          1 |     Total
          0 |        14         51 |        65
            |     21.54      78.46 |    100.00
          1 |        27         38 |        65
            |     41.54      58.46 |    100.00
      Total |        41         89 |       130
            |     31.54      68.46 |    100.00
```

```
. bysort tx: tabstat cr, stat(n mean) by(male) col(stat)
-> tx = 0
Summary for variables: cr
     by categories of: male
    male |         N      mean
---------+--------------------
       0 |        30        .7
       1 |        35  .4857143
---------+--------------------
   Total |        65  .5846154
---------+--------------------

-> tx = 1
Summary for variables: cr
     by categories of: male
    male |         N      mean
---------+--------------------
       0 |        35  .8571429
       1 |        30        .7
---------+--------------------
   Total |        65  .7846154
---------+--------------------
```

Problem 1b: Comparing the incidence of complete remission by treatment group
I choose to focus in the difference in probability of CR

```
. cs cr tx
                 | tx                     |
                 |   Exposed   Unexposed  |      Total
-----------------+------------------------+-----------
           Cases |        51          38  |         89
        Noncases |        14          27  |         41
-----------------+------------------------+-----------
           Total |        65          65  |        130
                 |                        |
            Risk |  .7846154    .5846154  |   .6846154
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
-----------------+------------------------+--------------------------
 Risk difference |                 .2     |    .0439899    .3560101
      Risk ratio |           1.342105     |    1.054389    1.708332
 Attr. frac. ex. |            .254902     |    .0515836    .4146336
 Attr. frac. pop |           .1460674     |
-----------------+------------------------+--------------------------
                               chi2(1) =     6.02  Pr>chi2 = 0.0141
```

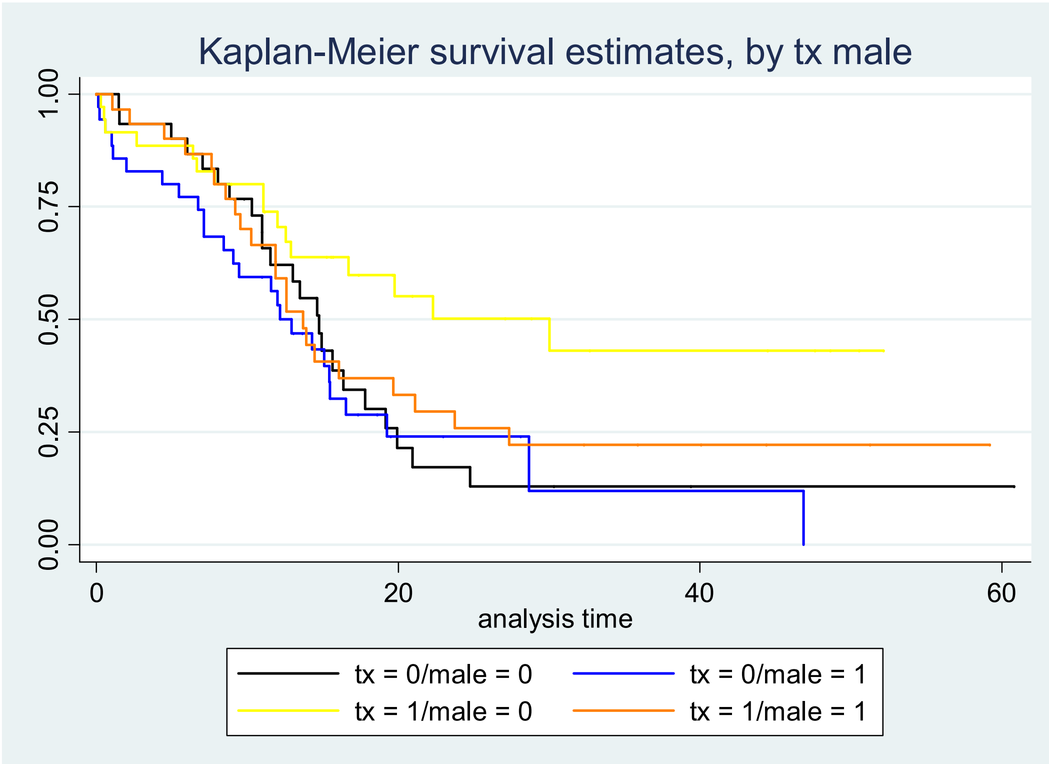
Problem 1c: Assessing the possibility of confounding by sex
Association between sex and treatment can be assessed from part 1a
Now looking at an association between CR and sex in the control group

```
. cs cr male if tx==0
                 | male                   |
                 |   Exposed   Unexposed  |      Total
-----------------+------------------------+-----------
```

```
                 Cases |        17            21  |         38
              Noncases |        18             9  |         27
-----------------------+--------------------------+-----------
                 Total |        35            30  |         65
                       |                          |
                  Risk |  .4857143            .7  |   .5846154
                       |                          |
                       |      Point estimate      |    [95% Conf. Interval]
-----------------------+--------------------------+------------------------
       Risk difference |        -.2142857         |  -.4473239    .0187525
            Risk ratio |         .6938776         |   .4588232     1.04935
       Prev. frac. ex. |         .3061224         |  -.0493498    .5411768
       Prev. frac. pop |         .1648352         |
                       +---------------------------------------------------
                               chi2(1) =     3.05  Pr>chi2 = 0.0805

#### Problem 1d: Comparing the incidence of complete remission by treatment group
#### after adjustment for sex using the Mantel-Haenszel statistic

. cs cr tx, by(male) or
                  male |       OR       [95% Conf. Interval]   M-H Weight
-----------------------+-------------------------------------------------
                     0 |    2.571429      .7807859   8.404801     1.615385 (Cornfield)
                     1 |    2.470588      .8981522   6.780976     2.353846 (Cornfield)
-----------------------+-------------------------------------------------
                 Crude |    2.588346       1.20652   5.545979
          M-H combined |    2.511628      1.144075   5.513866
-------------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =    0.002  Pr>chi2 = 0.9609

                   Test that combined OR = 1:
                                Mantel-Haenszel chi2(1) =      5.33
                                                Pr>chi2 =    0.0210

#### Problem 1e: Comparing the incidence of complete remission by treatment group
#### after adjustment for sex using logistic regression. I use the robust SE to
#### ensure that any lack of model fit does not lead to anti-conservative inference
#### (I could not be sure that the lack of model fit would necessarily have lead
#### to anti-conservative inference—it could have been conservative.)

. logistic cr tx male, robust
Logistic regression                               Number of obs   =        130
                                                  Wald chi2(2)    =      10.41
                                                  Prob > chi2     =     0.0055
Log pseudolikelihood = -75.252997                 Pseudo R2       =     0.0713
------------------------------------------------------------------------------
             |               Robust
          cr | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          tx |   2.511585   1.011183     2.29   0.022     1.140901    5.529019
        male |   .3981549   .1603001    -2.29   0.022     .1808639    .8765004
------------------------------------------------------------------------------
```

```
#### Problem 1g: Assessing the possibility of an interaction between treatment and sex
#### This could have been assessed using the output for part 1d. Here I use logistic regression
#### I do not use the robust standard error estimates in this "saturated" model in which
#### each of the four treatment-sex groups is estimated independently (there are four
#### parameters in the model)

. g txm= tx * male

. logistic cr tx male txm
Logistic regression                               Number of obs   =        130
                                                  LR chi2(3)      =      11.57
                                                  Prob > chi2     =     0.0090
Log likelihood = -75.251793                       Pseudo R2       =     0.0714
------------------------------------------------------------------------------
          cr | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          tx |   2.571429   1.610099     1.51   0.131     .7536936    8.77312
        male |   .4047619    .211528    -1.73   0.084     .1453326   1.127292
         txm |   .9607843   .7835976    -0.05   0.961     .1942683    4.75171
------------------------------------------------------------------------------


#####################################################################################
#### Problem 2: Analysis of time to death by treatment and sex
#####################################################################################

#### Problem 2a: Descriptive statistics
#### I prefer to measure time to death in months, rather than days

. replace ttodeath= ttodeath/30.4
(130 real changes made)

. stset ttodeath death
     failure event:  death != 0 & death < .
obs. time interval:  (0, ttodeath]
 exit on or before:  failure

      130  total obs.
        0  exclusions
---------------------------------------------------------
      130  obs. remaining, representing
       87  failures in single record/single failure data
    64669  total analysis time at risk, at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =      1848
```

```
. sts graph , by(tx male) col(black blue yellow orange)
         failure _d:  death
   analysis time _t:  ttodeath
```




```
. sts list, by(tx male) at(12 24)
         failure _d:  death
   analysis time _t:  ttodeath
              Beg.                             Survivor      Std.
    Time     Total      Fail                   Function     Error       [95% Conf. Int.]
-----------------------------------------------------------------------------------------
tx=0 male=0
      12        18        11                     0.6206    0.0907      0.4192     0.7697
      24         5        11                     0.1721    0.0767      0.0555     0.3427
tx=0 male=1
      12        18        16                     0.5317    0.0859      0.3526     0.6820
      24         4         8                     0.2406    0.0817      0.1030     0.4095
```

```
tx=1 male=0
      12        23        9              0.7385    0.0751     0.5570    0.8546
      24        11        6              0.5017    0.0968     0.3033    0.6712
tx=1 male=1
      12        17       12              0.5911    0.0913     0.3925    0.7441
      24         8        9              0.2586    0.0835     0.1154    0.4287
--------------------------------------------------------------------------------
Note:  Survivor function is calculated over full data and evaluated at
       indicated times; it is not calculated from aggregates shown at left.
```

Problem 2b: Analysis of a treatment effect using proportional hazards regression.
I use robust SE to relax assumptions about proportional hazards

```
. stcox tx, robust
         failure _d:  death
   analysis time _t:  ttodeath
Iteration 0:   log pseudolikelihood =   -368.044
Iteration 1:   log pseudolikelihood = -365.41537
Iteration 2:   log pseudolikelihood = -365.41523
Refining estimates:
Iteration 0:   log pseudolikelihood = -365.41523

Cox regression -- Breslow method for ties

No. of subjects      =          130                Number of obs    =       130
No. of failures      =           87
Time at risk         =  2127.269739
                                                   Wald chi2(1)     =      5.29
Log pseudolikelihood =   -365.41523                Prob > chi2      =    0.0214
--------------------------------------------------------------------------------
             |               Robust
          t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------------------------------------------------------------------------
         tx |     .60737   .1316259    -2.30   0.021     .3971784    .9287977
--------------------------------------------------------------------------------
```

Problem 2c: Assessing the possibility of confounding by sex
Relationships between sex and treatment were examined in problem 1a. Here I
consider the existence of an association between time to death and sex independent
of treatment by only using the control group.

```
. stcox male if tx==0, robust
         failure _d:  death
   analysis time _t:  ttodeath
Iteration 0:   log pseudolikelihood =  -166.9936
```

```
Iteration 1:   log pseudolikelihood = -166.84092
Iteration 2:   log pseudolikelihood = -166.84092
Refining estimates:
Iteration 0:   log pseudolikelihood = -166.84092
Cox regression -- Breslow method for ties
No. of subjects      =           65                Number of obs   =        65
No. of failures      =           49
Time at risk         =  899.9342122
                                                   Wald chi2(1)    =      0.32
Log pseudolikelihood =   -166.84092                Prob > chi2     =    0.5702
------------------------------------------------------------------------------
             |               Robust
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |   1.171496   .3265651     0.57   0.570     .6783604    2.023116
------------------------------------------------------------------------------
```

Problem 2d: Analysis of a treatment effect after adjustment for sex
using proportional hazards regression.
I use robust SE to relax assumptions about proportional hazards

```
. stcox tx male, robust
         failure _d:  death
   analysis time _t:  ttodeath
Iteration 0:   log pseudolikelihood =   -368.044
Iteration 1:   log pseudolikelihood = -364.12782
Iteration 2:   log pseudolikelihood = -364.12768
Refining estimates:
Iteration 0:   log pseudolikelihood = -364.12768
Cox regression -- Breslow method for ties
No. of subjects      =          130                Number of obs   =       130
No. of failures      =           87
Time at risk         =  2127.269739
                                                   Wald chi2(2)    =      7.60
Log pseudolikelihood =   -364.12768                Prob > chi2     =    0.0223
------------------------------------------------------------------------------
             |               Robust
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          tx |   .6182218   .1335386    -2.23   0.026     .4048378    .9440773
        male |   1.413015   .2972737     1.64   0.100     .9355541    2.134149
------------------------------------------------------------------------------
```

Problem 2f: Investigating the possibility of an interaction between treatment and
sex.
I use robust SE to relax assumptions about proportional hazards

```
. stcox tx male txm, robust
         failure _d:  death
   analysis time _t:  ttodeath
Iteration 0:   log pseudolikelihood =   -368.044
Iteration 1:   log pseudolikelihood = -363.73989
Iteration 2:   log pseudolikelihood = -363.65662
Iteration 3:   log pseudolikelihood = -363.65657
Refining estimates:
Iteration 0:   log pseudolikelihood = -363.65657
Cox regression -- Breslow method for ties
No. of subjects      =          130                Number of obs    =       130
No. of failures      =           87
Time at risk         =  2127.269739
                                                   Wald chi2(3)     =      7.45
Log pseudolikelihood =   -363.65657                Prob > chi2      =    0.0588
------------------------------------------------------------------------------
             |               Robust
          t  | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          tx |   .4898034   .1594606    -2.19   0.028     .2587643    .9271271
        male |   1.174754   .3115312     0.61   0.544     .6985833    1.975494
         txm |   1.525956   .6617335     0.97   0.330     .6522543    3.569991
------------------------------------------------------------------------------
```