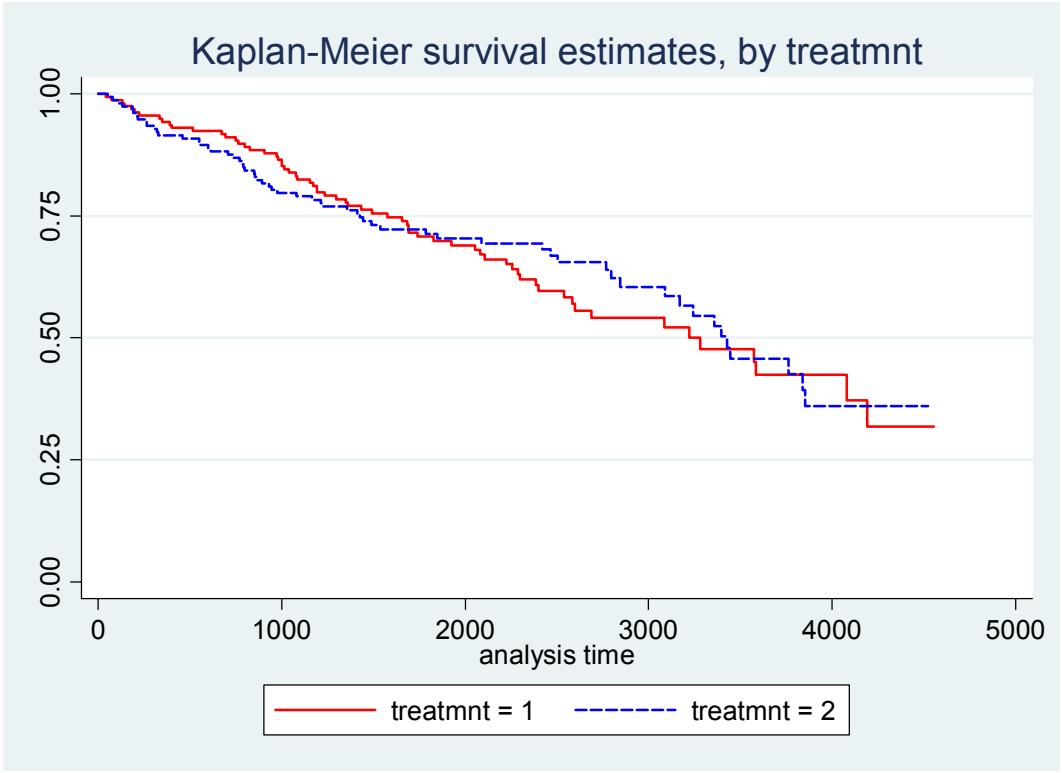
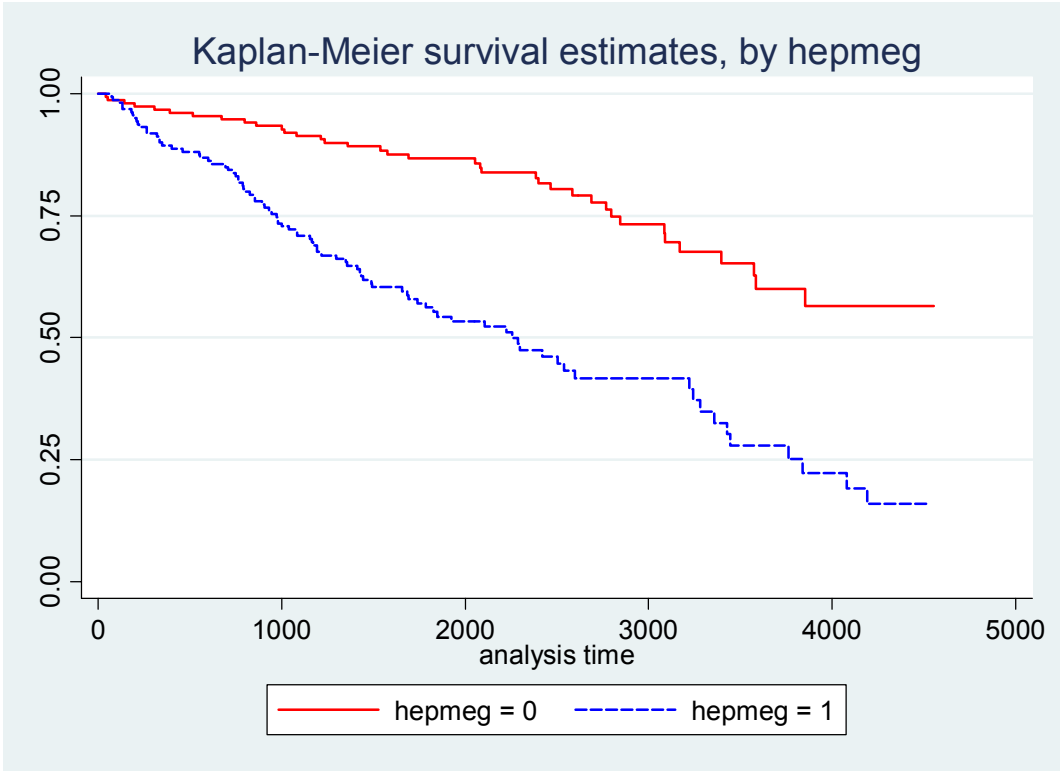
```
#### Biost 517: Applied Biostatistics I
#### Emerson, Fall 2005

#### Annotated Stata Log File: Homework #7
#### December 9, 2005

#### In this file I give the Stata commands I used to produce
#### the key to Homework #7. 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.

-------------------------------------------------------------------------------
#### Reading in the data: I had saved it previously

. use ..\..\datasets\2005.09.26\dfmowide


#### NOTE: Recall that Stata has some problem with representing the
####   dose 0.4 group correctly. Hence, I always selected that group
####   by asking for dose > 0.3.


#### Reading in the data; then saving it for the future
. infile age albumin alkphos ascites bili cholest edema edmadj hepmeg obstime
> platelet protime sex sgot spiders stage status treatmnt triglyc urinecu
> using liver.txt
#### (a whole bunch of warnings about missing data deleted)
 (419 observations read)

. drop in 1
(1 observation deleted)

. save liver
file liver.dta saved

#### Declaring the variables to be used in survival analyses
. stset obstime status

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

-------------------------------------------------------------------------------
      418  total obs.
        0  exclusions
-------------------------------------------------------------------------------
      418  obs. remaining, representing
      161  failures in single record/single failure data
   801633  total analysis time at risk, at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =      4795
```

```
####################
#### Problem 1: Comparing Kaplan-Meier estimates of survival probabilities
####          across treatment groups
####################

####   Generate Kaplan-Meier plots of survival probabilities by treatment
. sts graph, by(treatmnt) col("red" "blue") lp(l "-")

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




```
####   Generate listing of Kaplan-Meier estimates of survival probabilities at
####   5 and 10 years by treatment group
. sts list, by(treatmnt) at(1826 3652)

         failure _d:  status
   analysis time _t:  obstime

              Beg.                    Survivor      Std.
    Time     Total      Fail            Function     Error     [95% Conf. Int.]
-------------------------------------------------------------------------------
treatmnt=1
    1826        83        43              0.7071    0.0380      0.6252    0.7743
    3652        17        20              0.4244    0.0603      0.3054    0.5383
treatmnt=2
    1826        78        42              0.7134    0.0377      0.6319    0.7800
    3652        17        15              0.4567    0.0611      0.3344    0.5708
-------------------------------------------------------------------------------
Note:  Survivor function is calculated over full data and evaluated at
       indicated times; it is not calculated from aggregates shown at left.
```

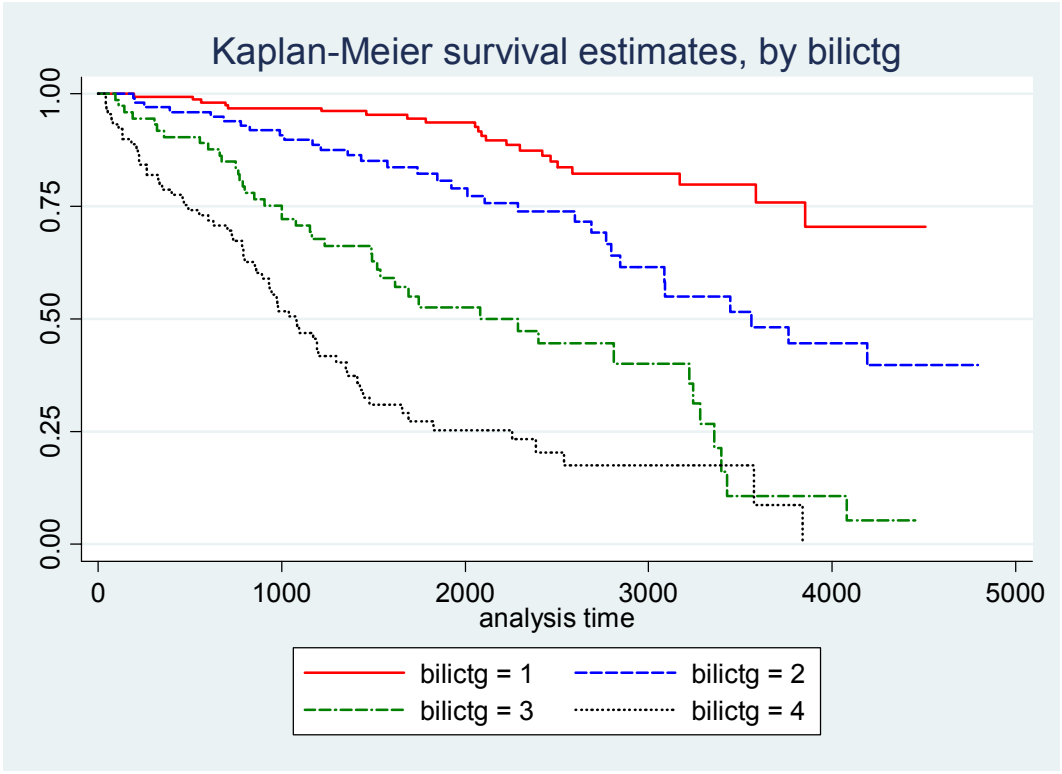
```
####   Computing difference (D-penicillamine - placebo) in 10 year survival probs
. display (0.4244 - 0.4567)
-.0323

####   Computing 95% CI for difference in 10 year survival probabilities
. display (0.4244 - 0.4567) - invnorm(.975) * sqrt(0.0603^2 + 0.0611^2)
-.20055238

. display (0.4244 - 0.4567) + invnorm(.975) * sqrt(0.0603^2 + 0.0611^2)
.13595238

####   Computing Z score to test for nonzero difference in 10 year survival
. display (0.4244 - 0.4567) / sqrt(0.0603^2 + 0.0611^2)
-.37626117

####   Computing two-sided P value to test for nonzero difference in 10 year survival
. display 2 * norm(-.37626117)
.70672274

####################
#### Problem 2: Comparing Kaplan-Meier estimates of survival probabilities
####         and hazard ratio across hepatomegaly groups
####################

####   Generate Kaplan-Meier plots of survival probabilities by hepatomegaly
. sts graph, by(hepmeg) col("red" "blue") lp(l "-")

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




```
####   Generate listing of Kaplan-Meier estimates of survival probabilities at
```

```
####   5 and 10 years by hepatomegaly
. sts list, by(hepmeg) at(1826 3652)

         failure _d:  status
   analysis time _t:  obstime

              Beg.                        Survivor      Std.
    Time     Total     Fail                Function     Error     [95% Conf. Int.]
-------------------------------------------------------------------------------
hepmeg=0
    1826        99       19                0.8670    0.0286     0.7989    0.9133
    3652        22       17                0.6000    0.0622     0.4676    0.7095
hepmeg=1
    1826        62       66                0.5615    0.0412     0.4769    0.6378
    3652        13       18                0.2786    0.0562     0.1753    0.3915
-------------------------------------------------------------------------------
Note:  Survivor function is calculated over full data and evaluated at
       indicated times; it is not calculated from aggregates shown at left.

####   Proportional hazards regression model to make inference about hazard ratio
####   (Note that I used robust standard error estimates)
. stcox hepmeg, robust

         failure _d:  status
   analysis time _t:  obstime

Iteration 0:   log pseudolikelihood = -639.97989
Iteration 1:   log pseudolikelihood = -619.94034
Iteration 2:   log pseudolikelihood = -619.92965
Iteration 3:   log pseudolikelihood = -619.92965
Refining estimates:
Iteration 0:   log pseudolikelihood = -619.92965

Cox regression -- Breslow method for ties

No. of subjects      =          312                 Number of obs    =       312
No. of failures      =          125
Time at risk         =       625985
                                                    Wald chi2(1)     =     37.24
Log pseudolikelihood =   -619.92965                 Prob > chi2      =    0.0000

------------------------------------------------------------------------------
             |               Robust
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      hepmeg |   3.274276   .6363758     6.10   0.000     2.237067    4.792384
------------------------------------------------------------------------------

####################
#### Problem 3: Comparing survival by bilirubin level at randomization.
####           For descriptive purposes, I create four categories of bilirubin levels
####           according to an approximate doubling of bilirubin levels. For inference,
####           I modeled bilirubin continuously untransformed. (In real life, I would
####           have log transformed bilirubin, so I also include that analysis for
####           comparison.
####################

####   Descriptive statistics for bilirubin just to use as a guide when creating the
```

```
####   bilirubin strata.

. tabstat bili, stat(n mean sd min p25 p50 p75 max) col(stat)

variable |   N      mean        sd   min       p25       p50       p75       max
---------+--------------------------------------------------------------------
    bili | 418  3.220813  4.407506    .3        .8       1.4       3.4        28
------------------------------------------------------------------------------

####   Generating the categorized bilirubin variable.
. g bilictg=bili

. recode bilictg 0/1=1 1/2=2 2/4=3 4/max=4
(bilictg: 391 changes made)

. tabstat bili, stat(n mean sd min p25 p50 p75 max) col(stat) by(bilictg)

Summary for variables: bili
     by categories of: bilictg

 bilictg |   N      mean        sd    min       p25       p50       p75       max
---------+--------------------------------------------------------------------
       1 | 157  .6917197  .1779375     .3        .6        .7        .8         1
       2 |  99  1.433333  .2910221    1.1       1.2       1.3       1.7         2
       3 |  73  2.906849   .567482    2.1       2.3         3       3.4         4
       4 |  89   9.92809  5.572497    4.2       5.6       7.3      13.8        28
---------+--------------------------------------------------------------------
   Total | 418  3.220813  4.407506     .3        .8       1.4       3.4        28
------------------------------------------------------------------------------


####   Generate Kaplan-Meier plots of survival probabilities by bilirubin strata
. sts graph, by(bilictg) col("red" "blue" "green" "black") lp(l "-" "--." ".")

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




```
####   Generate listing of Kaplan-Meier estimates of survival probabilities at
####   5 and 10 years by bilirubin strata
. sts list, by(bilictg) at(1826 3652)

         failure _d:  status
   analysis time _t:  obstime

              Beg.                    Survivor      Std.
    Time     Total      Fail           Function     Error     [95% Conf. Int.]
-------------------------------------------------------------------------------
bilictg=1
    1826       108         9             0.9363    0.0207      0.8805    0.9666
    3652        19        12             0.7588    0.0573      0.6239    0.8509
bilictg=2
    1826        54        16             0.8219    0.0410      0.7242    0.8877
    3652        15        14             0.4813    0.0786      0.3221    0.6238
bilictg=3
    1826        24        31             0.5265    0.0645      0.3934    0.6432
    3652         3        10             0.1070    0.0669      0.0209    0.2749
bilictg=4
    1826        15        59             0.2734    0.0529      0.1759    0.3799
    3652         2         5             0.0873    0.0672      0.0108    0.2688
-------------------------------------------------------------------------------
Note:  Survivor function is calculated over full data and evaluated at
       indicated times; it is not calculated from aggregates shown at left.

####   Proportional hazards regression model to make inference about hazard ratio
####   when fitting bilirubin continuously and untransformed.
####   (Note that I used robust standard error estimates)
. stcox bili, robust
```

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

Iteration 0:   log pseudolikelihood = -873.48766
Iteration 1:   log pseudolikelihood = -832.65308
Iteration 2:   log pseudolikelihood = -824.89161
Iteration 3:   log pseudolikelihood = -824.82202
Iteration 4:   log pseudolikelihood =   -824.822
Refining estimates:
Iteration 0:   log pseudolikelihood =   -824.822

Cox regression -- Breslow method for ties

No. of subjects      =          418                 Number of obs    =       418
No. of failures      =          161
Time at risk         =       801633
                                                    Wald chi2(1)     =    100.63
Log pseudolikelihood =     -824.822                 Prob > chi2      =    0.0000

------------------------------------------------------------------------------
             |               Robust
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       bili |   1.152364   .0162909    10.03   0.000     1.120872     1.18474
------------------------------------------------------------------------------

####   Proportional hazards regression model to make inference about hazard ratio
####   when fitting bilirubin continuously as a log transformed variable.
####   (Note that I used robust standard error estimates)
. g logbili= log(bili)

. stcox logbili, robust

        failure _d:  status
  analysis time _t:  obstime

Iteration 0:   log pseudolikelihood = -873.48766
Iteration 1:   log pseudolikelihood = -803.14827
Iteration 2:   log pseudolikelihood = -796.99011
Iteration 3:   log pseudolikelihood = -796.97982
Refining estimates:
Iteration 0:   log pseudolikelihood = -796.97982

Cox regression -- Breslow method for ties

No. of subjects      =          418                 Number of obs    =       418
No. of failures      =          161
Time at risk         =       801633
                                                    Wald chi2(1)     =    153.51
Log pseudolikelihood =   -796.97982                 Prob > chi2      =    0.0000

------------------------------------------------------------------------------
             |               Robust
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     logbili |   2.687895   .2145033    12.39   0.000     2.298707    3.142977
```

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

####   Transforming parameter estimates to reflect doubling of bilirubin
. display 2.687895^log(2)
1.9844763

. display 2.298707^log(2)
1.7805794

. display 3.142977^log(2)
2.2117226
```