Chapter 10 Stata v10.1 Analysis Examples Syntax and Output

General Notes on Stata 10.1

Given that this tool is used throughout the ASDA textbook this chapter includes only the syntax and output for the analysis examples provided in Chapter 10. Stata 10.1 is an excellent tool for survey data analysis as well as graphing and related data management tasks. It offers a very comprehensive set of svy commands as well as weighted graphics and convenient syntax and data management abilities. For these reasons, we use Stata as the primary software for the ASDA text.

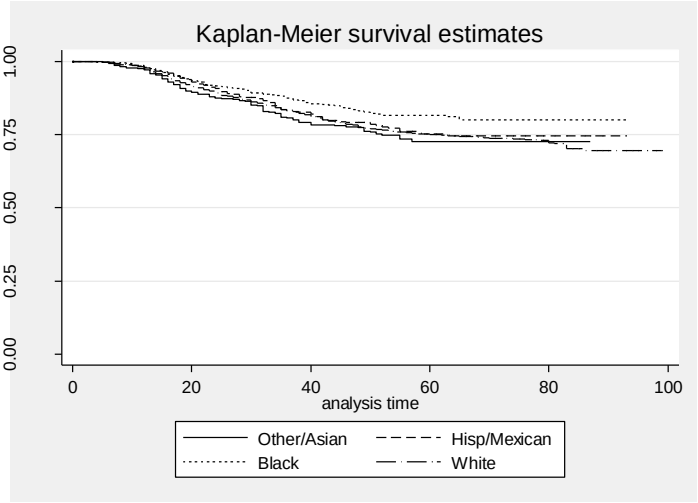
The examples and syntax presented here assume that all data management including variable construction, labels for variable values and other preparation steps are complete. See the Stata documentation for assistance with these issues.

Nearly all analysis examples presented can be done in Stata 10.1 and are included in this chapter's output. The exception is complex sample survey adjusted standard errors for the Kaplan-Meier survival curves which are presented from Sudaan in the ASDA textbook.

Please check the Stata documentation and also the ASDA web site for updates to Stata as new versions are released. For example, we have already included an example of how to use Stata 11.0 with the new "factor" variable features/syntax and compared this to the older "xi" type of syntax for including categorical variables in data analysis.

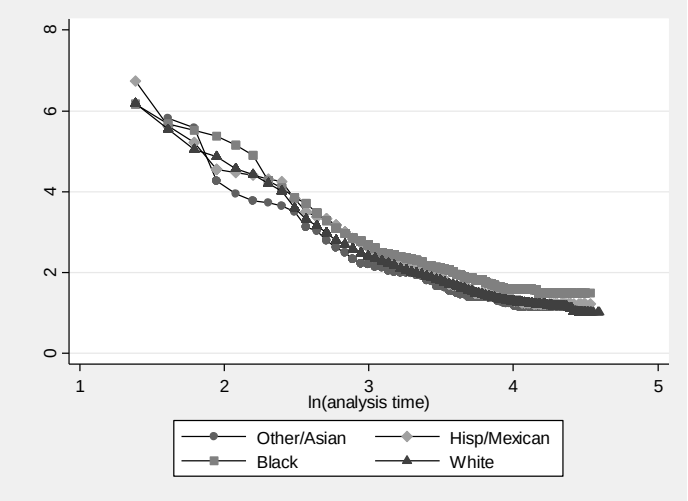
CHAPTER 10 SURVIVAL ANALYSIS STATA 10 NCS-R DATA

```
. * figure 10.3
. stset ageonsetmde [pweight=ncsrwtsh] , failure(mde==1)

     failure event:  mde == 1
obs. time interval:  (0, ageonsetmde]
 exit on or before:  failure
           weight:  [pweight=ncsrwtsh]

--------------------------------------------------------------------------------
     9282  total obs.
        0  exclusions
--------------------------------------------------------------------------------
     9282  obs. remaining, representing
     1829  failures in single record/single failure data
   385696  total analysis time at risk, at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =        99

. label var racecat "1=Other 2=His/Mex 3=AA/AfCar 4=White"

. * graphs by race groups using the stset with weight applied
. sts graph, by(racecat) legend(lab(1 "Other/Asian") lab(2 "Hisp/Mexican") lab(3 "Black") lab(4
"White"))

         failure _d:  mde == 1
   analysis time _t:  ageonsetmde
             weight:  [pweight=ncsrwtsh]

.
```



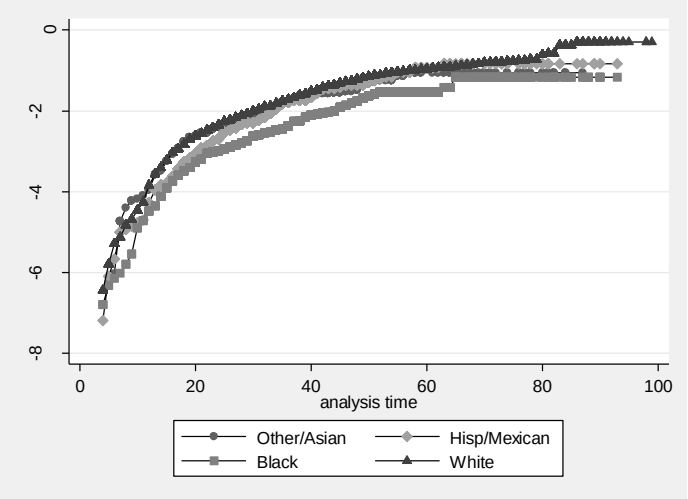

```
.
. * add the list command for sts
. sts list , at(10 20 30 40 50 60 70)

         failure _d:  mde == 1
   analysis time _t:  ageonsetmde
             weight:  [pweight=ncsrwtsh]

           Beg.                       Survivor
    Time     Total     Fail             Function
-------------------------------------------------
      10   9171.84  139.004              0.9850
      20   8169.95  632.016              0.9165
      30   6390.66  401.574              0.8659
      40   4736.49  327.113              0.8155
      50   2997.27  186.745              0.7760
      60   1779.61  59.0606              0.7568
      70   949.309  22.2811              0.7438
-------------------------------------------------
Note:  survivor function is calculated over full data and evaluated at
       indicated times; it is not calculated from aggregates shown at left.
```

```
. * figure 10.4 : test the PH assumption for race groups NCS-R DATA

. stphplot , by(racecat) legend(lab(1 "Other/Asian") lab(2 "Hisp/Mexican") lab(3 "Black") lab(4
"White"))

        failure _d:  mde == 1
  analysis time _t:  ageonsetmde
            weight:  [pweight=ncsrwtsh]
```




```
. * Table 10.2 Cox Model NCS-R DATA

. svyset seclustr [pweight=ncsrwtsh], strata(sestrat) vce(linearized) singleunit(missing)

      pweight: ncsrwtsh
          VCE: linearized
  Single unit: missing
     Strata 1: sestrat
         SU 1: seclustr
        FPC 1: <zero>

. gen intwage=age

. char sex[omit] 2

. xi: svy linearized : stcox age i.sex i.mar3cat i.ed4cat i.racecat
i.sex              _Isex_1-2           (naturally coded; _Isex_2 omitted)
i.mar3cat          _Imar3cat_1-3       (naturally coded; _Imar3cat_1 omitted)
i.ed4cat           _Ied4cat_1-4        (naturally coded; _Ied4cat_1 omitted)
i.racecat          _Iracecat_1-4       (naturally coded; _Iracecat_1 omitted)
(running stcox on estimation sample)

Survey: Cox regression

Number of strata   =        42                  Number of obs      =      9282
Number of PSUs     =        84                  Population size    = 9282.0002
                                                Design df          =        42
                                                F(  10,     33)    =     53.02
                                                Prob > F           =    0.0000

------------------------------------------------------------------------------
             |             Linearized
          _t | Haz. Ratio   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .9517007   .0022648   -20.80   0.000      .947141    .9562823
     _Isex_1 |   .6353491   .0395875    -7.28   0.000     .5602771    .7204801
 _Imar3cat_2 |   1.652846   .0992196     8.37   0.000     1.464266    1.865713
 _Imar3cat_3 |   1.083971   .0962944     0.91   0.369      .906064    1.296809
  _Ied4cat_2 |   .9447047   .0633411    -0.85   0.401     .8251482    1.081584
  _Ied4cat_3 |   1.046723   .0607273     0.79   0.436     .9310729    1.176738
  _Ied4cat_4 |   .9137495   .0581822    -1.42   0.164      .803564    1.039044
 _Iracecat_2 |   .7788393   .1046541    -1.86   0.070     .5938527     1.02145
 _Iracecat_3 |   .6193833   .0922551    -3.22   0.003     .4585813    .8365707
 _Iracecat_4 |   1.080887   .1269548     0.66   0.511      .852783    1.370005
------------------------------------------------------------------------------
```

```
. * figure 10.5 test proportional hazards assumption for the fitted Cox model
. *alternative transformation for the PH assumption and adjusted for other covariates in the model
. stphplot, by(racecat) adj(age sexf swd nevermarried ed12 ed1315 ed16 ) nonegative nolntime ///
> legend(lab(1 "Other/Asian") lab(2 "Hisp/Mexican") lab(3 "Black") lab(4 "White"))

        failure _d:  mde == 1
  analysis time _t:  ageonsetmde
            weight:  [pweight=ncsrwtsh]
```



```
. * Output for 10.5.5 : show some data records for understanding of how key variables are created
. list  caseid intwage ncsrwtsh sestrat seclustr pyr mdetv ageonsetmde if caseid ==1

     +-------------------------------------------------------------------------+
     | caseid   intwage   ncsrwtsh   sestrat   seclustr   pyr   mdetv   ageons~e |
     |-------------------------------------------------------------------------|
  1. |      1        41    2.02426         1          2     1       0         34 |
  2. |      1        41    2.02426         1          2     2       0         34 |
  3. |      1        41    2.02426         1          2     3       0         34 |
  4. |      1        41    2.02426         1          2     4       0         34 |
  5. |      1        41    2.02426         1          2     5       0         34 |
     |-------------------------------------------------------------------------|
  6. |      1        41    2.02426         1          2     6       0         34 |
  7. |      1        41    2.02426         1          2     7       0         34 |
  8. |      1        41    2.02426         1          2     8       0         34 |
  9. |      1        41    2.02426         1          2     9       0         34 |
 10. |      1        41    2.02426         1          2    10       0         34 |
     |-------------------------------------------------------------------------|
 11. |      1        41    2.02426         1          2    11       0         34 |
 12. |      1        41    2.02426         1          2    12       0         34 |
 13. |      1        41    2.02426         1          2    13       0         34 |
 14. |      1        41    2.02426         1          2    14       0         34 |
 15. |      1        41    2.02426         1          2    15       0         34 |
     |-------------------------------------------------------------------------|
 16. |      1        41    2.02426         1          2    16       0         34 |
 17. |      1        41    2.02426         1          2    17       0         34 |
 18. |      1        41    2.02426         1          2    18       0         34 |
 19. |      1        41    2.02426         1          2    19       0         34 |
 20. |      1        41    2.02426         1          2    20       0         34 |
     |-------------------------------------------------------------------------|
 21. |      1        41    2.02426         1          2    21       0         34 |
 22. |      1        41    2.02426         1          2    22       0         34 |
 23. |      1        41    2.02426         1          2    23       0         34 |
 24. |      1        41    2.02426         1          2    24       0         34 |
 25. |      1        41    2.02426         1          2    25       0         34 |
     |-------------------------------------------------------------------------|
 26. |      1        41    2.02426         1          2    26       0         34 |
 27. |      1        41    2.02426         1          2    27       0         34 |
 28. |      1        41    2.02426         1          2    28       0         34 |
 29. |      1        41    2.02426         1          2    29       0         34 |
 30. |      1        41    2.02426         1          2    30       0         34 |
     |-------------------------------------------------------------------------|
 31. |      1        41    2.02426         1          2    31       0         34 |
 32. |      1        41    2.02426         1          2    32       0         34 |
 33. |      1        41    2.02426         1          2    33       0         34 |
 34. |      1        41    2.02426         1          2    34       1         34 |
 35. |      1        41    2.02426         1          2    35       0         34 |
     |-------------------------------------------------------------------------|
 36. |      1        41    2.02426         1          2    36       0         34 |
 37. |      1        41    2.02426         1          2    37       0         34 |
 38. |      1        41    2.02426         1          2    38       0         34 |
 39. |      1        41    2.02426         1          2    39       0         34 |
 40. |      1        41    2.02426         1          2    40       0         34 |
     |-------------------------------------------------------------------------|
 41. |      1        41    2.02426         1          2    41       0         34 |
     +-------------------------------------------------------------------------+
```

```
* table 10.5 discrete time logistic regression NCS-R DATA (PERSONYEAR DATA STRUCTURE)

. char sex[omit] 2

. xi: svy linearized : logistic mdetv pyr age i.sex i.ed4cat i.racecat I.mar3cat  if int <= ageonsetmde
i.sex             _Isex_1-2           (naturally coded; _Isex_2 omitted)
i.ed4cat          _Ied4cat_1-4        (naturally coded; _Ied4cat_1 omitted)
i.racecat         _Iracecat_1-4       (naturally coded; _Iracecat_1 omitted)
I.mar3cat         _Imar3cat_1-3       (naturally coded; _Imar3cat_1 omitted)
(running logistic on estimation sample)

Survey: Logistic regression

Number of strata   =        42                  Number of obs      =    385696
Number of PSUs     =        84                  Population size    = 386866.05
                                                Design df          =        42
                                                F(  11,     32)    =     53.63
                                                Prob > F           =    0.0000

------------------------------------------------------------------------------
             |             Linearized
       mdetv | Odds Ratio   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         pyr |   1.033342   .0021429    15.82   0.000     1.029026    1.037675
         age |   .9433351   .0023099   -23.82   0.000     .9386851    .9480081
     _Isex_1 |   .6409081    .039921    -7.14   0.000     .5652022    .7267544
  _Ied4cat_2 |   .9800651   .0647966    -0.30   0.762     .8576485    1.119955
  _Ied4cat_3 |   1.097373   .0630381     1.62   0.113     .9772536    1.232256
  _Ied4cat_4 |    .980737   .0621179    -0.31   0.760     .8630591     1.11446
 _Iracecat_2 |   .7800309   .1051252    -1.84   0.072     .5942829    1.023836
 _Iracecat_3 |   .6332007   .0949096    -3.05   0.004     .4679213    .8568604
 _Iracecat_4 |   1.076802   .1273204     0.63   0.535      .848215    1.366991
 _Imar3cat_2 |   1.639269   .1000123     8.10   0.000     1.449366    1.854053
 _Imar3cat_3 |   .9652712   .0849149    -0.40   0.690     .8082557    1.152789
------------------------------------------------------------------------------

. xi: svy: logistic, coef

Survey: Logistic regression

Number of strata   =        42                  Number of obs      =    385696
Number of PSUs     =        84                  Population size    = 386866.05
                                                Design df          =        42
                                                F(  11,     32)    =     53.63
                                                Prob > F           =    0.0000

------------------------------------------------------------------------------
             |             Linearized
       mdetv |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         pyr |   .0327978   .0020738    15.82   0.000     .0286128    .0369828
         age |  -.0583337   .0024486   -23.82   0.000    -.0632752   -.0533922
     _Isex_1 |  -.4448692   .0622881    -7.14   0.000    -.5705718   -.3191667
  _Ied4cat_2 |  -.0201363   .0661146    -0.30   0.762     -.153561    .1132884
  _Ied4cat_3 |   .0929188   .0574446     1.62   0.113    -.0230091    .2088466
  _Ied4cat_4 |   -.019451   .0633379    -0.31   0.760    -.1472721    .1083702
 _Iracecat_2 |  -.2484217   .1347706    -1.84   0.072    -.5203997    .0235564
 _Iracecat_3 |  -.4569678   .1498886    -3.05   0.004    -.7594553   -.1544803
 _Iracecat_4 |   .0739955   .1182393     0.63   0.535    -.1646211    .3126122
 _Imar3cat_2 |   .4942501   .0610103     8.10   0.000     .3711263     .617374
 _Imar3cat_3 |  -.0353462     .08797    -0.40   0.690    -.2128768    .1421844
       _cons |  -3.435525   .1619878   -21.21   0.000    -3.762429    -3.10862
------------------------------------------------------------------------------
```

```
. * table 10.6 CLOGLOG FOR COMPARISON TO LOGIT

. char sex[omit] 2

. xi: svy linearized : cloglog mdetv pyr age i.sex i.ed4cat i.racecat I.mar3cat  if int <= ageonsetmde
i.sex             _Isex_1-2           (naturally coded; _Isex_2 omitted)
i.ed4cat          _Ied4cat_1-4        (naturally coded; _Ied4cat_1 omitted)
i.racecat         _Iracecat_1-4       (naturally coded; _Iracecat_1 omitted)
I.mar3cat         _Imar3cat_1-3       (naturally coded; _Imar3cat_1 omitted)
(running cloglog on estimation sample)

Survey: Complementary log-log regression

Number of strata   =        42                  Number of obs      =    385696
Number of PSUs     =        84                  Population size    = 386866.05
                                                Design df          =        42
                                                F(  11,     32)    =     53.65
                                                Prob > F           =    0.0000

------------------------------------------------------------------------------
             |             Linearized
       mdetv |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         pyr |   .0327329    .002069    15.82   0.000     .0285575    .0369083
         age |  -.0581804   .0024412   -23.83   0.000    -.0631069   -.0532538
     _Isex_1 |  -.4432213   .0620831    -7.14   0.000      -.56851   -.3179326
  _Ied4cat_2 |  -.0197399   .0658767    -0.30   0.766    -.1526845    .1132047
  _Ied4cat_3 |   .0923596   .0572221     1.61   0.114    -.0231193    .2078385
  _Ied4cat_4 |  -.0192036   .0631122    -0.30   0.762    -.1465693     .108162
 _Iracecat_2 |  -.2474242   .1343775    -1.84   0.073    -.5186091    .0237606
 _Iracecat_3 |  -.4550781   .1494743    -3.04   0.004    -.7567295   -.1534268
 _Iracecat_4 |   .0737345   .1178773     0.63   0.535    -.1641514    .3116205
 _Imar3cat_2 |   .4928149   .0607862     8.11   0.000     .3701433    .6154865
 _Imar3cat_3 |  -.0354735   .0875323    -0.41   0.687    -.2121209    .1411739
       _cons |  -3.444394   .1613723   -21.34   0.000    -3.770056   -3.118731
------------------------------------------------------------------------------

. xi: svy linearized : cloglog mdetv pyr age i.sex i.ed4cat i.racecat I.mar3cat  if int <= ageonsetmde,
eform
i.sex             _Isex_1-2           (naturally coded; _Isex_2 omitted)
i.ed4cat          _Ied4cat_1-4        (naturally coded; _Ied4cat_1 omitted)
i.racecat         _Iracecat_1-4       (naturally coded; _Iracecat_1 omitted)
I.mar3cat         _Imar3cat_1-3       (naturally coded; _Imar3cat_1 omitted)
(running cloglog on estimation sample)

Survey: Complementary log-log regression

Number of strata   =        42                  Number of obs      =    385696
Number of PSUs     =        84                  Population size    = 386866.05
                                                Design df          =        42
                                                F(  11,     32)    =     53.65
                                                Prob > F           =    0.0000

------------------------------------------------------------------------------
             |             Linearized
       mdetv |     exp(b)   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         pyr |   1.033275   .0021378    15.82   0.000     1.028969    1.037598
         age |   .9434798   .0023032   -23.83   0.000     .9388431    .9481393
     _Isex_1 |   .6419651   .0398552    -7.14   0.000     .5663687    .7276518
  _Ied4cat_2 |   .9804536   .0645891    -0.30   0.766     .8584005    1.119861
  _Ied4cat_3 |   1.096759   .0627589     1.61   0.114     .9771459    1.231014
  _Ied4cat_4 |   .9809796   .0619118    -0.30   0.762     .8636659    1.114228
 _Iracecat_2 |   .7808094   .1049232    -1.84   0.073      .595348    1.024045
 _Iracecat_3 |   .6343984   .0948263    -3.04   0.004     .4691984    .8577636
 _Iracecat_4 |   1.076521   .1268974     0.63   0.535     .8486135    1.365636
 _Imar3cat_2 |   1.636917   .0995021     8.11   0.000     1.447942    1.850557
 _Imar3cat_3 |   .9651483   .0844817    -0.41   0.687     .8088669    1.151625
------------------------------------------------------------------------------
```