Chapter 11 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 11. 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.

All analysis examples presented can be done in Stata 10.1 and are included in this chapter's output.

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 11 MISSING DATA STATA 10

```
*HRS data examine missing data patterns for some key variables
*table 11.1
. * examine missing data patterns in some selected variables in HRS 2006 data
. mvpatterns bodywgt totheight kage gender racecat arthritis diabetes numfalls24 if age65p==1
variables with no mv's: kage gender

Variable     | type     obs   mv   variable label
-------------+-------------------------------------
bodywgt      | float  11562  169
totheight    | float  11566  165
racecat      | byte   11730    1
arthritis    | float  11713   18
diabetes     | byte   11717   14
numfalls24   | float  11197  534
-------------------------------------------------

Patterns of missing values

  +-------------------------+
  | _pattern   _mv   _freq |
  |-------------------------|
  |   ++++++     0   10891 |
  |   +++++.     1     508 |
  |   +.++++     1     130 |
  |   .+++++     1     123 |
  |   ..++++     2      27 |
  |-------------------------|
  |   .++++.     2      16 |
  |   +++.++     1      13 |
  |   ++++.+     1       9 |
  |   +.+++.     2       5 |
  |   ++++..     2       2 |
  |-------------------------|
  |   +++..+     2       2 |
  |   +++.+.     2       1 |
  |   +..+++     2       1 |
  |   .++.++     2       1 |
  |   ..+++.     3       1 |
  |-------------------------|
  |   ..+...     5       1 |
  +-------------------------+
```

```
tab numfalls24 if age65p==1, missing

 numfalls24 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      7,507       63.99       63.99
          1 |      1,616       13.78       77.77
          2 |        924        7.88       85.64
          3 |        450        3.84       89.48
          4 |        239        2.04       91.52
          5 |        137        1.17       92.69
          6 |        107        0.91       93.60
          7 |         24        0.20       93.80
          8 |         24        0.20       94.01
          9 |          2        0.02       94.02
         10 |         61        0.52       94.54
         11 |          5        0.04       94.59
         12 |         41        0.35       94.94
         13 |          1        0.01       94.95
         15 |         12        0.10       95.05
         20 |         23        0.20       95.24
         22 |          1        0.01       95.25
         24 |          6        0.05       95.30
         25 |          5        0.04       95.35
         30 |          2        0.02       95.36
         40 |          4        0.03       95.40
         48 |          1        0.01       95.41
         50 |          5        0.04       95.45
          . |        534        4.55      100.00
------------+-----------------------------------
      Total |     11,731      100.00


. sum kage if age65p==1, detail

                    age at 2006 interview
-------------------------------------------------------------
      Percentiles      Smallest
 1%            65            65
 5%            65            65
10%            66            65       Obs                11731
25%            69            65       Sum of Wgt.        11731

50%            74                     Mean           75.14091
                        Largest       Std. Dev.      7.716989
75%            80           102
90%            86           103       Variance       59.55192
95%            89           103       Skewness       .6978065
99%            95           105       Kurtosis       2.757613

. tab gender if age65p==1, missing

     gender |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      5,002       42.64       42.64
          2 |      6,729       57.36      100.00
------------+-----------------------------------
      Total |     11,731      100.00

. tab arthritis if age65p==1, missing

  arthritis |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      3,736       31.85       31.85
          1 |      7,977       68.00       99.85
          . |         18        0.15      100.00
------------+-----------------------------------
      Total |     11,731      100.00

. tab diabetes if age65p==1, missing

   diabetes |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      9,103       77.60       77.60
          1 |      2,614       22.28       99.88
          . |         14        0.12      100.00
------------+-----------------------------------
      Total |     11,731      100.00
```

```
. sum bodywgt if age65p==1, detail

                          bodywgt
-------------------------------------------------------------
      Percentiles      Smallest
 1%           98             65
 5%          113             68
10%          122             73       Obs                11562
25%          140             75       Sum of Wgt.        11562

50%          166                      Mean           169.8123
                        Largest       Std. Dev.      39.55376
75%          194            350
90%          220            360       Variance         1564.5
95%          240            365       Skewness       .6670401
99%          285            380       Kurtosis       3.815928

. sum totheight if age65p==1, detail

                         totheight
-------------------------------------------------------------
      Percentiles      Smallest
 1%           58             48
 5%           60             48
10%           61             48       Obs                11566
25%           63             48       Sum of Wgt.        11566

50%           66                      Mean           66.04332
                        Largest       Std. Dev.      4.098245
75%           69             82
90%           72             83       Variance       16.79562
95%           73             83       Skewness       .1200914
99%           75             84       Kurtosis       2.917842
```

```
*NOTE: NHANES DATA (ADULTS THAT COMPLETED MEDICAL EXAMINATION AND NON ZERO ON MEDICAL EXAM WEIGHT:
DEFINED BY : if age18p==1 & wtmec2yr > 0)



. * table 11.2 missing data patterns
. mvpatterns bpxdi1_1 marcat riagendr ridreth1 agec agecsq bmxbmi indfmpir if age18p==1 , nodrop
Variable     | type     obs   mv   variable label
-------------+-------------------------------------------------------------
bpxdi1_1     | int     4581  753
marcat       | byte    5329    5   1=married 2=prev married 3=never married
riagendr     | byte    5334    0   gender - adjudicated
ridreth1     | byte    5334    0   1=mex 2=oth hisp 3=white 4=black 5=other
agec         | float   5334    0
agecsq       | float   5334    0
bmxbmi       | float   5237   97   body mass index (kg/m**2)
indfmpir     | float   5066  268   family pir
-------------------------------------------------------------------------------

Patterns of missing values

  +------------------------+
  | _pattern   _mv   _freq |
  |------------------------|
  | ++++++++     0    4308 |
  | .+++++++     1     666 |
  | +++++++.     1     217 |
  | ++++++.+     1      49 |
  | .++++++.     2      41 |
  |------------------------|
  | .+++++.+     2      39 |
  | .+++++..     3       5 |
  | ++++++..     2       4 |
  | +.++++++     1       3 |
  | ..++++++     2       1 |
  |------------------------|
  | ..+++++.     3       1 |
  +------------------------+



. * Section 11.7.3 imputation examples
. * repeat run for Stata 10 examples uses different data set name
. ice bpxdi1_1 marcat m2 m3 r2 r3 r4 r5  g2 agec agecsq bmxbmi indfmpir  , ///
> saving(f:\applied_analysis_book\imputed_nhanes_example_feb2010) m(5) seed(123) replace  ///
> passive(m2:marcat==2\ m3:marcat==3) substitute(marcat: m2 m3)

   #missing |
     values |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      4,308       80.76       80.76
          1 |        935       17.53       98.29
          2 |         85        1.59       99.89
          3 |          6        0.11      100.00
------------+-----------------------------------
      Total |      5,334      100.00

   Variable | Command | Prediction equation
------------+---------+-------------------------------------------------------
   bpxdi1_1 | regress | m2 m3 r2 r3 r4 r5 g2 agec agecsq bmxbmi indfmpir
     marcat | mlogit  | bpxdi1_1 r2 r3 r4 r5 g2 agec agecsq bmxbmi indfmpir
         m2 |         | [Passively imputed from marcat==2]
         m3 |         | [Passively imputed from marcat==3]
         r2 |         | [No missing data in estimation sample]
         r3 |         | [No missing data in estimation sample]
         r4 |         | [No missing data in estimation sample]
         r5 |         | [No missing data in estimation sample]
         g2 |         | [No missing data in estimation sample]
       agec |         | [No missing data in estimation sample]
     agecsq |         | [No missing data in estimation sample]
     bmxbmi | regress | bpxdi1_1 m2 m3 r2 r3 r4 r5 g2 agec agecsq indfmpir
   indfmpir | regress | bpxdi1_1 m2 m3 r2 r3 r4 r5 g2 agec agecsq bmxbmi
-------------------------------------------------------------------------------

Imputing ..........1.........2.........3.........4.........5
(note: file f:\applied_analysis_book\imputed_nhanes_example_feb2010.dta not found)
file f:\applied_analysis_book\imputed_nhanes_example_feb2010.dta saved
```

```
. * use imputed feb 2010 file and do analysis needed
. use "f:\applied_analysis_book\imputed_nhanes_example_feb2010.dta" , clear
. mean bmxbmi indfmpir bpxdi1_1 [pweight=wtmec2yr] , over(_mj)

Mean estimation                     Number of obs    =   30981

            0: _mj = 0
            1: _mj = 1
            2: _mj = 2
            3: _mj = 3
            4: _mj = 4
            5: _mj = 5

--------------------------------------------------------------
        Over |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
bmxbmi       |
           0 |   28.34348   .1255658      28.09737     28.5896
           1 |   28.43542   .1157619      28.20852    28.66232
           2 |   28.42367   .1158188      28.19666    28.65068
           3 |   28.44852   .1155165       28.2221    28.67494
           4 |   28.42029   .1159863      28.19295    28.64763
           5 |   28.44997   .1155902      28.22341    28.67653
-------------+------------------------------------------------
indfmpir     |
           0 |   3.111066   .0291669      3.053898    3.168234
           1 |   3.057951    .026597       3.00582    3.110082
           2 |   3.060913   .0265877        3.0088    3.113026
           3 |   3.065391    .026532      3.013387    3.117395
           4 |   3.058774   .0265924      3.006652    3.110896
           5 |   3.067349   .0266414      3.015131    3.119568
-------------+------------------------------------------------
bpxdi1_1     |
           0 |   70.71618   .2288887      70.26755    71.16481
           1 |   70.49577   .2079759      70.08813    70.90341
           2 |   70.48846   .2081999      70.08038    70.89654
           3 |   70.47491   .2108867      70.06156    70.88825
           4 |   70.50781   .2097281      70.09673    70.91889
           5 |    70.3497   .2121438      69.93389    70.76551
--------------------------------------------------------------
. proportion marcat [pweight=wtmec2yr] , over(_mj)

Proportion estimation               Number of obs    =   31999

      Married: marcat = Married
      _prop_2: marcat = Previously Married
      _prop_3: marcat = Never Married

            0: _mj = 0
            1: _mj = 1
            2: _mj = 2
            3: _mj = 3
            4: _mj = 4
            5: _mj = 5

--------------------------------------------------------------
        Over | Proportion   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
Married      |
           0 |    .639239   .0080863      .6233895    .6550885
           1 |   .6390962    .008084      .6232512    .6549412
           2 |    .639004   .0080833      .6231605    .6548475
           3 |   .6394344   .0080795      .6235984    .6552704
           4 |    .639245   .0080804      .6234071    .6550829
           5 |   .6390184   .0080831      .6231751    .6548617
-------------+------------------------------------------------
_prop_2      |
           0 |   .1824983   .0064451      .1698656     .195131
           1 |   .1828985   .0064516      .1702532    .1955438
           2 |   .1828013   .0064471      .1701648    .1954377
           3 |   .1825603   .0064399      .1699379    .1951827
           4 |   .1825603   .0064399      .1699379    .1951827
           5 |   .1829763   .0064487      .1703366     .195616
-------------+------------------------------------------------
_prop_3      |
           0 |   .1782627   .0061993      .1661119    .1904136
           1 |   .1780053   .0061913      .1658701    .1901405
           2 |   .1781947   .0061934      .1660553    .1903341
           3 |   .1780053   .0061913      .1658701    .1901405
           4 |   .1781947   .0061934      .1660553    .1903341
           5 |   .1780053   .0061913      .1658701    .1901405
--------------------------------------------------------------
```

```
. * table 11.4
. svyset sdmvpsu [pweight=wtmec2yr], strata(sdmvstra)

      pweight: wtmec2yr
          VCE: linearized
  Single unit: missing
     Strata 1: sdmvstra
         SU 1: sdmvpsu
        FPC 1: <zero>

. *mim: mean bpxdi1_1
.
. mim: svy: mean bpxdi1_1

Multiple-imputation estimates (svy: mean)                Imputations =       5
Survey: Mean estimation                                  Minimum obs =    5334
                                                         Minimum dof =    12.5

------------------------------------------------------------------------------
             |     Coef.  Std. Err.     t    P>|t|    [95% Conf. Int.]   MI.df
-------------+----------------------------------------------------------------
    bpxdi1_1 |   70.4633   .311088  226.51   0.000    69.7887  71.1379    12.5
------------------------------------------------------------------------------

*imputation variance statistics

. matrix list e(MIM_W)

symmetric e(MIM_W)[1,1]
           bpxdi1_1
bpxdi1_1  .09176358

. matrix list e(MIM_B)

symmetric e(MIM_B)[1,1]
           bpxdi1_1
bpxdi1_1  .00417692

. matrix list e(MIM_T)

symmetric e(MIM_T)[1,1]
           bpxdi1_1
bpxdi1_1  .09677588

. matrix list e(MIM_dfvec)

symmetric e(MIM_dfvec)[1,1]
           c1
r1  12.536471
```

```
*table 11.5
. * use mim with svy for full regression model
. xi: mim: svy: regress bpxdi1_1 i.marcat i.riagendr i.ridreth1 agec agecsq bmxbmi indfmpir
i.marcat          _Imarcat_1-3        (naturally coded; _Imarcat_1 omitted)
i.riagendr        _Iriagendr_1-2      (naturally coded; _Iriagendr_1 omitted)
i.ridreth1        _Iridreth1_1-5      (naturally coded; _Iridreth1_1 omitted)

Multiple-imputation estimates (svy: regress)             Imputations =       5
Survey: Linear regression                                Minimum obs =    5334
                                                         Minimum dof =     7.3

------------------------------------------------------------------------------
    bpxdi1_1 |     Coef.  Std. Err.     t    P>|t|    [95% Conf. Int.]   MI.df
-------------+----------------------------------------------------------------
  _Imarcat_2 |   .799518   .700033    1.14   0.287   -.819931  2.41897     7.9
  _Imarcat_3 |   -.40967   .656801   -0.62   0.552   -1.95085  1.13152     7.3
_Iriagendr_2 |  -2.77956   .422086   -6.59   0.000   -3.73188 -1.82725     9.2
_Iridreth1_2 |   1.17618   1.02832    1.14   0.277   -1.08556  3.43792    11.1
_Iridreth1_3 |   2.10938   .599131    3.52   0.005     .78116  3.43759    10.4
_Iridreth1_4 |   3.13496   .793606    3.95   0.003    1.34227  4.92766     9.1
_Iridreth1_5 |     1.915   .930508    2.06   0.062   -.116479  3.94647    11.8
        agec |    .12134   .013316    9.11   0.000    .092179  .150501    11.5
      agecsq |  -.012037    .00086  -13.99   0.000   -.014043  -.01003     7.5
      bmxbmi |   .183199   .033057    5.54   0.000    .111437  .254961    12.4
    indfmpir |  -.126685    .15096   -0.84   0.424   -.469608  .216237     8.8
       _cons |   68.6909   1.22081   56.27   0.000    66.0263  71.3555    11.8
------------------------------------------------------------------------------


. mim: testparm _Imarcat_2 _Imarcat_3 _Iridreth1_2 _Iridreth1_3 _Iridreth1_4 _Iridreth1_5

 ( 1)  _Imarcat_2 = 0
 ( 2)  _Imarcat_3 = 0
 ( 3)  _Iridreth1_2 = 0
 ( 4)  _Iridreth1_3 = 0
 ( 5)  _Iridreth1_4 = 0
 ( 6)  _Iridreth1_5 = 0

       F(  6, 301.9) =    4.39
            Prob > F =    0.0003
```

```
*compare to svy regress with complete case analysis approach

use "f:\applied_analysis_book\impute_subset_nhanes.dta" if age18p==1 & wtmec2yr > 0

. svyset sdmvpsu [pweight=wtmec2yr], strata(sdmvstra)

      pweight: wtmec2yr
          VCE: linearized
  Single unit: missing
     Strata 1: sdmvstra
         SU 1: sdmvpsu
        FPC 1: <zero>

. xi: svy: regress bpxdi1_1 i.marcat i.riagendr i.ridreth1 agec agecsq bmxbmi indfmpir
i.marcat          _Imarcat_1-3        (naturally coded; _Imarcat_1 omitted)
i.riagendr        _Iriagendr_1-2      (naturally coded; _Iriagendr_1 omitted)
i.ridreth1        _Iridreth1_1-5      (naturally coded; _Iridreth1_1 omitted)
(running regress on estimation sample)

Survey: Linear regression

Number of strata   =        15                 Number of obs      =       4308
Number of PSUs     =        30                 Population size    =  181480256
                                                Design df          =         15
                                                F(  11,      5)    =     118.59
                                                Prob > F           =     0.0000
                                                R-squared          =     0.1491

------------------------------------------------------------------------------
             |             Linearized
    bpxdi1_1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  _Imarcat_2 |    1.03395   .6933599     1.49   0.157     -.443912    2.511811
  _Imarcat_3 |   -.372906   .5731209    -0.65   0.525    -1.594484    .8486722
_Iriagendr_2 |  -2.699262   .3888311    -6.94   0.000    -3.528036   -1.870488
_Iridreth1_2 |   1.652752   1.160156     1.42   0.175    -.8200616    4.125566
_Iridreth1_3 |   2.145357   .6611924     3.24   0.005     .7360593    3.554656
_Iridreth1_4 |   3.310581    .868765     3.81   0.002     1.458852     5.16231
_Iridreth1_5 |   1.663891     .92322     1.80   0.092    -.3039057    3.631688
        agec |   .1193541   .0156409     7.63   0.000     .0860163    .1526918
      agecsq |  -.0123063   .0008857   -13.89   0.000    -.0141941   -.0104185
      bmxbmi |   .1973282    .038926     5.07   0.000     .1143593     .280297
    indfmpir |  -.1076662   .1376078    -0.78   0.446    -.4009702    .1856379
       _cons |   68.28714   1.451376    47.05   0.000      65.1936    71.38067
------------------------------------------------------------------------------
```