APPENDIX F
EXAMPLES IN STATA

In Stata, the NCVS sample design must be appropriately specified using the *svyset* command, as follows, each time a new data set is put in use for analysis. Note that the nested stratification requires creation of a stratum recode prior to *svyset*.

```
egen strat=group(yr_grp v2117)
svyset v2118 [pweight=newwgt], strata(strat) vce(linearized)
```

Note that examples are presented as if they were run in sequence. Therefore, data sets are cleared and set to use and the *svyset* command is included only when necessary and not for each example.

For the calculation of victimization rates, analysts must first create a new variable equal to the product of the victimization count and the adjustment factor (*ADJINC_WT*), multiplied by 1,000 (as outlined in the examples below). The *subpop* option is used to limit the analysis to the appropriate set of cases. Prior to each analysis a "sub" variable is created based on all exclusions (e.g. victimizations occurring outside of the US and outside the year(s) of interest).

Examples 1 and *2* demonstrate how to estimate the total number of victimizations. *Examples 3* and *4* demonstrate how to calculate the proportion of victimizations with given characteristics. *Examples 5* and *6* demonstrate the calculation of personal and property victimization rates for victimization characteristics included on the modified person and household files, and *Example 7* demonstrates the calculation of victimization rates for victimization characteristics not on the modified files. Examples are included for both single- and pooled-year estimates. Finally, *Example 8* demonstrates how to identify low-quality estimates and implement the rounding recommended as discussed in ***Section 4*** of the user's guide.

F.1 Victimization Totals

The Stata *total* command is used to estimate the total number of victimizations from the modified incident-level file. The domain(s) of interest (i.e. subsets of the population based on characteristics of the victimization or victim) are specified with the *over* option, and the subpopulation is specified within the *svy* prefix command, which must be included for each

analytic command. *Examples 1A-1C* demonstrate the calculation of victimization totals for single years, and *Examples 2A-2C* demonstrate pooled year estimates.

Example 1: Number of victimizations, single year

Examples 1A and *1B* below are estimates of personal crimes, while *Example 1C* is an estimate of property crimes. *Examples 1A* and *1C* provide overall crime estimates, while *Example 1B* provides estimated totals for person-level characteristics of interest.

Example 1A - Total number of violent victimizations, 2011

Year(s): 2011
Crime Type: violent victimizations (*vcrime*)
Domain(s): n/a
Weight: *newwgt*
Subpopulation*: wit*hin the United States (*exclude_outUS*=**0**); non-dummy records (*dummy*=**0**); *2011 (year=**2011**)*

Code:

```
use "<long path of incident-level .dta file>" //a
egen strat=group(yr_grp v2117) //b
svyset v2118 [pweight=newwgt], strata(strat) vce(linearized) //c

generate sub=(exclude_outus==0 & dummy==0 & year==2011) //d
svy, subpop(sub): total vcrime //e
```

Code Comment(s):

a) Read in input data
b) Create nested stratification recode
c) Specify the NCVS design
d) Create subpopulation indicator
e) Estimate totals

Output:

```
Survey: Total estimation

Number of strata =     160       Number of obs    =      37853
Number of PSUs   =     320       Population size  =  115824878
                                 Subpop. no. obs  =       7255
                                 Subpop. size     = 23041440.8
                                 Design df        =        160

--------------------------------------------------------------
             |             Linearized
             |      Total   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
      vcrime |    5812523   357912.4       5105681     6519364
--------------------------------------------------------------
```

Example 1B - Total number of violent victimizations by sex, age category, and race/ethnicity, 2011

Year(s): *2011*
Crime Type: *violen*t victimizations (*vcrime*)
Domain(s): sex (*sex*), age category (*agecat*), and race/ethnicity (*race_eth*)
Weight: *newwgt*
Subpopulation: *wit*hin the United States (*exclude_outUS*=**0**); non-dummy records (*dummy*=**0**); *2011 (year=**2011**)*

Code:

```
svy, subpop(sub): total vcrime, over(sex)
svy, subpop(sub): total vcrime, over(agecat)
svy, subpop(sub): total vcrime, over(race_eth)
```

Output:

```
Survey: Total estimation

Number of strata =     160       Number of obs    =      37853
Number of PSUs   =     320       Population size  =  115824878
                                 Subpop. no. obs  =       7255
                                 Subpop. size     = 23041440.8
                                 Design df        =        160

         Male: sex = Male
       Female: sex = Female

------------------------------------------------------------------
             |             Linearized
        Over |      Total   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------
vcrime       |
        Male |    3209725   236960.6       2741751     3677699
      Female |    2602798   252785.9       2103571     3102025
------------------------------------------------------------------
```

```
Survey: Total estimation

Number of strata =     160       Number of obs    =      37853
Number of PSUs   =     320       Population size  =  115824878
                                 Subpop. no. obs  =       7255
                                 Subpop. size     = 23041440.8
                                 Design df        =        160

    _subpop_1: agecat = 12 to 14
    _subpop_2: agecat = 15 to 17
    _subpop_3: agecat = 18 to 20
    _subpop_4: agecat = 21 to 24
    _subpop_5: agecat = 25 to 34
    _subpop_6: agecat = 35 to 49
    _subpop_7: agecat = 50 to 64
    _subpop_8: agecat = 65 or older

------------------------------------------------------------------
             |             Linearized
        Over |      Total   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------
vcrime       |
   _subpop_1 |   488382.9   77367.96      335588.8      641177
   _subpop_2 |   428879.6    64692.5      301118.3    556640.9
   _subpop_3 |   843753.1   220103.6      409070.3     1278436
   _subpop_4 |   617719.6   69925.52      479623.6    755815.6
   _subpop_5 |    1114834   118262.1        881278     1348390
   _subpop_6 |    1365450   141010.7       1086968     1643933
   _subpop_7 |   776857.5   90922.04      597295.4    956419.5
   _subpop_8 |   176645.7   46321.04      85166.22    268125.2
------------------------------------------------------------------
```

```
Survey: Total estimation

Number of strata =     160       Number of obs    =      37853
Number of PSUs   =     320       Population size  =  115824878
                                 Subpop. no. obs  =       7255
                                 Subpop. size     = 23041440.8
                                 Design df        =        160

    _subpop_1: race_eth = Non-Hispanic White
    _subpop_2: race_eth = Non-Hispanic Black
    _subpop_3: race_eth = Hispanic
    _subpop_4: race_eth = Non-Hispanic Other
    _subpop_5: race_eth = Non-Hispanic More than One Race

------------------------------------------------------------------
             |             Linearized
        Over |      Total   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------
vcrime       |
   _subpop_1 |    3719729   242039.2       3241725     4197733
   _subpop_2 |   812173.2   181536.1      453657.2     1170689
   _subpop_3 |     895434   130311.4      638081.9     1152786
   _subpop_4 |   197081.8   46736.13      104782.5    289381.1
   _subpop_5 |   188104.6   43090.36      103005.3    273203.8
------------------------------------------------------------------
```

Example 1C - Total number of property victimizations, 2011

Year(s): 2011
Crime Type: property victimizations (*pcrime*)
Domain(s): n/a
Weight: *newwgt*
Subpopulation: within the United States (*exclude_outUS*=**0**); non-dummy records (*dummy*=**0**); 2011 (*year*=**2011**)

Code:

```
svy, subpop(sub): total pcrime
display %20.10f _b[pcrime]
```

Output:

```
Survey: Total estimation

Number of strata =     160       Number of obs    =      37853
Number of PSUs   =     320       Population size  =  115824878
                                 Subpop. no. obs  =       7255
                                 Subpop. size     = 23041440.8
                                 Design df        =        160

--------------------------------------------------------------
             |             Linearized
             |      Total   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
      pcrime |   1.71e+07   543578.6      1.60e+07    1.81e+07
--------------------------------------------------------------

. display %20.10f _b[pcrime]
         17063147.83
```

Example 2: Number of crimes, aggregated years

Examples 2A and *2B* are estimates of personal crimes, while *Example 2C* is an estimate of a property crime. *Examples 2A* and *2C* provide overall crime estimates and *Example 2B* provides estimated totals for a victimization-level characteristic of interest. The code provided calculates the estimated number of victimization across the pooled years. To calculate the average number of victimizations per year, estimated totals and standard errors must be divided by the number of pooled years, as shown in the following examples.

Example 2A - Total and average number of aggravated assaults, 2009-2011

Year(s): 2009-2011
Crime Type: aggravated assaults (*aast*)
Domain(s): n/a
Weight: *newwgt*
Subpopulation: within the United States (*exclude_outUS*=**0**); non-dummy records (*dummy*=**0**); 2009-2011 ((**2009** <= *year*) and (*year* <=**2011**))

Code:

```
replace sub=(exclude_outus==0 & dummy==0 & year>=2009 ///
     & year<=2011) //a
svy, subpop(sub): total aast
```

Code Comment(s):

a) Subpopulation indicator incorporates the range of years to be analyzed

Output:

```
Survey: Total estimation

Number of strata =     160       Number of obs    =      37853
Number of PSUs   =     320       Population size  =  115824878
                                 Subpop. no. obs  =      20701
                                 Subpop. size     =   65552798
                                 Design df        =        160

--------------------------------------------------------------
             |             Linearized
             |      Total   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
        aast |    2940416   160814.1       2622823     3258008
--------------------------------------------------------------
```

NOTE: The estimate above represents the total number of aggravated assaults from 2009-2011. To obtain the average number of aggravated assaults, both the estimate and the standard error must be divided by the number of pooled years (3), as follows:

$$avg\ number\ of\ aggravated\ assaults = \frac{total\ number\ of\ aggravated\ asaults}{number\ of\ pooled\ years}$$

$$= \frac{2940416}{3}$$

$$= 980138.67$$

$$SE(avg\ number\ of\ aggravated\ assaults) = \frac{se(total\ number\ of\ aggravated\ asaults)}{number\ of\ pooled\ years}$$

$$= \frac{160814.1}{3}$$

$$= 53604.7$$

Example 2B - Total and average number of aggravated assaults involving firearm, 2009–2011

Year(s): 2009-2011
Crime Type: aggravated assaults (*aast*)
Domain(s): weapon category (*weapcat*)
Weight: *newwgt*
Subpopulation: within the United States (*exclude_outUS*=**0**); non-dummy records (*dummy*=**0**); 2009-2011 ((**2009** <= *year*) and (*year* <=**2011**))

Code:

```
svy, subpop(sub): total aast, over(weapcat)
```

Output:

```
Survey: Total estimation

Number of strata =     160        Number of obs    =     21040
Number of PSUs   =     320        Population size  =  67127145
                                  Subpop. no. obs  =      3888
                                  Subpop. size     =  16855065
                                  Design df        =       160

    _subpop_1: weapcat = No Weapon
    _subpop_2: weapcat = Firearm
    _subpop_3: weapcat = Knife or sharp object
    _subpop_4: weapcat = Other type weapon
    _subpop_5: weapcat = Type weapon unknown
    _subpop_6: weapcat = Do Not Know if off had weapon

--------------------------------------------------------------
             |             Linearized
        Over |      Total   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
aast         |
   _subpop_1 |   318541.9   53739.51      212411.7    424672.2
   _subpop_2 |   778706.3   81552.16      617648.8    939763.8
   _subpop_3 |   642022.3   65520.62      512625.6    771419.1
   _subpop_4 |   939927.3   96628.22      749096.1     1130759
   _subpop_5 |   228754.8   52287.67      125491.8    332017.8
   _subpop_6 |   32462.99   11505.77      9740.229    55185.74
--------------------------------------------------------------
```

NOTE: The estimate above represents the total number of aggravated assaults involving a firearm from 2009-2011. To obtain the average number of aggravated assaults involving a firearm, both the estimate and the standard error must be divided by the number of pooled years (3), as follows:

$$avg\ number\ of\ aggravated\ assaults\ w\ firearm = \frac{total\ number\ of\ aggravated\ asaults\ w\ firearm}{number\ of\ pooled\ years}$$

$$= \frac{778706.3}{3}$$

$$= 259568.77$$

$$SE(avg\ number\ of\ aggravated\ assaults\ w\ firearm) = \frac{se(total\ number\ of\ aggravated\ asaults\ w\ firearm)}{number\ of\ pooled\ years}$$

$$= \frac{81552.16}{3}$$

$$= 27184.05$$

Example 2C - Total and average number of household burglaries, 2009-2011

Year(s): 2009-2011
Crime Type: household burglary (*hburg*)
Domain(s): n/a
Weight: *newwgt*
Subpopulation: Within the United States (*exclude_outUS*=**0**); non-dummy records (*dummy*=**0**); 2009-2011 ((**2009** <= *year*) and (*year* <=**2011**))

Code:

```
svy, subpop(sub): total hburg
```

Output:

```
Survey: Total estimation

Number of strata =     160       Number of obs    =      37853
Number of PSUs   =     320       Population size  =  115824878
                                 Subpop. no. obs  =      20701
                                 Subpop. size     =   65552798
                                 Design df        =        160
--------------------------------------------------------------
             |             Linearized
             |      Total   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
       hburg |   1.02e+07   336711.5       9536345    1.09e+07
--------------------------------------------------------------

. display %20.2f _b[hburg]
         10201317.31
```

NOTE: The estimate above represents the total number of household burglaries from 2009-2011. To obtain the average number of household burglaries, both the estimate and the standard error must be divided by the number of pooled years (3), as follows:

$$avg\ number\ of\ household\ burglaries = \frac{total\ number\ of\ household\ burglaries}{number\ of\ pooled\ years}$$

$$= \frac{10201317.31}{3}$$

$$= 3400439.10$$

$$SE(avg\ number\ of\ household\ burglaries) = \frac{se(total\ number\ of\ household\ burglaries)}{number\ of\ pooled\ years}$$

$$= \frac{336711.5}{3}$$

$$= 112237.17$$

F.2 Victimization Proportions

The Stata *prop* command is used to estimate the percent distribution of victimizations in a specific domain across covariates from the modified incident-level file.

Example 3: Percent distribution of victims across characteristics, single year

Example 3A is an estimate of personal crimes, while *Example 3B* is an estimate of property crimes. In *Example 3A*, both the analysis domain and covariates are specified based on person- and incident-level characteristics. In Example 3B, the analysis domain is specified based on an incident-level characteristic, while the covariates are household-level characteristics.

Example 3A - Distribution of female violent crime victims by age category, race/ethnicity, and victim-offender relationship, 2011

Year(s): 2011
Domain(s): female (*sex*=2); violent victimizations (*vcrime*=1)
Covariate(s): age category (*agecat*); race/ethnicity (*race_eth*); victim-offender relationship (*direl*)
Weight: *newwgt*
Subpopulation: within the United States (*exclude_outUS*=**0**); non-dummy records (*dummy*=**0**); 2011 (*year*=**2011**); female (*sex*=2); violent victimizations (*vcrime*=1)

Code:

```
replace sub=(exclude_outus==0 & dummy==0 & year==2011 & /// sex==2 &
     vcrime==1) //a
svy, subpop(sub): prop agecat race_eth direl
```

Code Comment(s):

```
   a) Subpopulation indicator
```

Output:

```
Survey: Proportion estimation

Number of strata =     152       Number of obs    =      36478
Number of PSUs   =     304       Population size  =  111842311
                                 Subpop. no. obs  =        642
                                 Subpop. size     =  2602797.9
                                 Design df        =        152
```

```
      _prop_1: agecat = 12 to 14
      _prop_2: agecat = 15 to 17
      _prop_3: agecat = 18 to 20
      _prop_4: agecat = 21 to 24
```

```
      _prop_5: agecat = 25 to 34
      _prop_6: agecat = 35 to 49
      _prop_7: agecat = 50 to 64
      _prop_8: agecat = 65 or older
      _prop_9: race_eth = Non-Hispanic White
     _prop_10: race_eth = Non-Hispanic Black
     _prop_12: race_eth = Non-Hispanic Other
     _prop_13: race_eth = Non-Hispanic More than One Race
     _prop_15: direl = other relatives
     _prop_16: direl = well known/casual acquaintances
     _prop_18: direl = do not know relationship
     _prop_19: direl = do not know number of offenders

--------------------------------------------------------------
             |             Linearized
             | Proportion   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
agecat       |
     _prop_1 |   .0764283   .0185879      .0468795    .1222141
     _prop_2 |   .0626202   .0133092      .0409354    .0946582
     _prop_3 |   .1550198   .0536768      .0754829    .2919037
     _prop_4 |   .0931161   .0163669       .065431    .1308748
     _prop_5 |    .193023   .0236158      .1505845    .2439863
     _prop_6 |    .245841   .0305423      .1905589     .310999
     _prop_7 |   .1291292   .0199286       .094593    .1738535
     _prop_8 |   .0448224   .0172203       .020758    .0941036
-------------+------------------------------------------------
race_eth     |
     _prop_9 |   .5986556   .0445676      .5083471    .6827289
    _prop_10 |    .194289   .0518122       .111422    .3168124
    _prop_11 |   .1167428   .0198261      .0829049    .1619528
    _prop_12 |   .0442116   .0150832      .0223406    .0856186
    _prop_13 |    .046101   .0128716      .0263906    .0793332
-------------+------------------------------------------------

direl        |
    _prop_14 |   .2323057   .0350176      .1703326    .3084441
    _prop_15 |   .0998217   .0164185      .0717442    .1372627
    _prop_16 |   .3528307   .0455939      .2687222    .4471669
    _prop_17 |   .2533559    .031567      .1961575    .3205815
    _prop_18 |   .0270556   .0060991      .0172898    .0421011
    _prop_19 |   .0346305   .0093803      .0201909    .0587772
--------------------------------------------------------------
```

Example 3B - Distribution of property crime victims by household income, MSA status, and region, 2011

Year(s): 2011
Domain(s): property victimizations (*pcrime*=1)
Covariate(s): household income (*hincome*); MSA status (*msa*); region (*region*)
Weight: *newwgt*
Subpopulation: within the United States (*exclude_outUS*=**0**); non-dummy records (*dummy*=**0**); 2011 (*year*=**2011**); property victimizations (*pcrime*=1)

Code:

```
replace sub=(exclude_outus==0 & dummy==0 & year==2011 ///       &
      pcrime==1)
svy, subpop(sub): prop hincome msa region
```

Output:

```
Survey: Proportion estimation

Number of strata =     160       Number of obs    =      37853
Number of PSUs   =     320       Population size  =  115824878
                                 Subpop. no. obs  =       5857
                                 Subpop. size     = 17063147.8
                                 Design df        =        160

      _prop_1: hincome = Less than $7,500
      _prop_2: hincome = $7,500 to $14,999
      _prop_3: hincome = $15,000 to $24,999
      _prop_4: hincome = $25,000 to $34,999
      _prop_5: hincome = $35,000 to $49,999
      _prop_6: hincome = $50,000 to $74,999
      _prop_7: hincome = $75,000 or more
      _prop_8: hincome = Unknown
```

```
--------------------------------------------------------------
             |            Linearized
             | Proportion   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
hincome      |
     _prop_1 |   .0645022   .0067548      .0523787    .0791972
     _prop_2 |   .0867627   .0065909      .0745994    .1006934
     _prop_3 |   .1024865   .0065767      .0902078    .1162232
```

```
     _prop_4 |   .0888547   .0060064      .0776856    .1014531
     _prop_5 |   .1171942    .006256      .1053927    .1301249
     _prop_6 |   .1149773   .0069595      .1019313    .1294523
     _prop_7 |   .1646692   .0073005      .1507533    .1795979
     _prop_8 |   .2605532   .0112839      .2388927    .2834463
-------------+------------------------------------------------
msa          |
       Urban |   .4059674   .0151111       .376505    .4361227
    Suburban |   .4661668   .0161604      .4344328     .498177
       Rural |   .1278657   .0162809      .0990085    .1636065
-------------+------------------------------------------------
region       |
   Northeast |   .1340047   .0108131        .11405    .1568326
     Midwest |   .2141045   .0144778      .1868994     .244081
       South |   .3480723   .0151951      .3186965    .3786511
        West |   .3038185    .013962      .2769696    .3320746
-------------------------------------------------------------
```

Example 4: Percent distribution of victims across characteristics, aggregated years

Example 4A is an estimate of personal crimes, while *Example 4B* is an estimate of property crimes. Both examples specify multiple covariates based on incident-level characteristics.

Example 4A - Percent of violent victimizations reported and not reported to police by type of crime, 2009-2011

Year(s): 2009-2011
Domain(s): violent victimizations (*vcrime*=1)
Covariate(s): report status (*notify*); type of crime (*newoff*)
Weight: *newwgt*
Subpopulation: within the United States (*exclude_outUS*=**0**); non-dummy records (*dummy*=**0**); 2009-2011 ((**2009** <= *year*) and (*year* <=**2011**)); violent victimizations (*vcrime*=1)

Code:

```
replace sub=(exclude_outus==0 & dummy==0 & year>=2009 ///       &
      year<=2011 & vcrime==1)
svy, subpop(sub): prop notify, over(newoff)
```

Output:

```
Survey: Proportion estimation

Number of strata =     160       Number of obs    =      37837
Number of PSUs   =     320       Population size  =  115775395
                                 Subpop. no. obs  =       3744
                                 Subpop. size     = 16368259.8
                                 Design df        =        160

      _prop_1: notify = reported to police
      _prop_2: notify = not reported to police
      _prop_3: notify = do not know

    _subpop_1: newoff = Rape & Sexual Assault
    _subpop_2: newoff = Robbery
    _subpop_3: newoff = Aggravated Assault
    _subpop_4: newoff = Simple Assault
```

	Linearized			
Over	Proportion	Std. Err.	[95% Conf.	Interval]
_prop_1				
_subpop_1	.324803	.0641144	.2126291	.4614713
_subpop_2	.6159029	.0332389	.5485241	.6791087
_subpop_3	.590134	.0255435	.5389107	.6394737
_subpop_4	.4412306	.017906	.4062232	.4768322
_prop_2				
_subpop_1	.675197	.0641144	.5385287	.7873709
_subpop_2	.3765311	.0338498	.3123699	.4453358
_subpop_3	.3949489	.0266856	.3436439	.4486774
_subpop_4	.544175	.0179034	.5086541	.5792521
_prop_3				
_subpop_1	.	(no observations)		
_subpop_2	.007566	.0046406	.0022445	.0251858
_subpop_3	.0149171	.0093124	.0043131	.0502748
_subpop_4	.0145944	.0044527	.0079715	.0265724

Example 4B - Percent of property victimizations reported and not reported to police by type of crime, 2009-2011

Year(s): 2009-2011
Domain(s): property victimizations (*pcrime*=1)
Covariate(s): report status (*notify*); type of crime (*newoff*)
Weight: *newwgt*
Subpopulation: within the United States (*exclude_outUS*=**0**); non-dummy records (*dummy*=**0**); 2009-2011 ((**2009** <= *year*) and (*year* <=**2011**)); property victimizations (*pcrime*=1)

Code:

```
replace sub=(exclude_outus==0 & dummy==0 & year>=2009 ///
      & year<=2011 & pcrime==1)
svy, subpop(sub): prop notify, over(newoff)
```

Output:

```
Survey: Proportion estimation

Number of strata =     160       Number of obs    =      37824
Number of PSUs   =     320       Population size  =  115743034
                                 Subpop. no. obs  =      16784
                                 Subpop. size     = 48615889.5
                                 Design df        =        160

      _prop_1: notify = reported to police
      _prop_2: notify = not reported to police
      _prop_3: notify = do not know

    _subpop_1: newoff = Household Burglary
    _subpop_2: newoff = Motor Vehicle Theft
    _subpop_3: newoff = Theft


--------------------------------------------------------------
             |             Linearized
        Over | Proportion   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
_prop_1      |
   _subpop_1 |   .5455355   .0150821      .5156222    .5751237
   _subpop_2 |   .8373771   .0139997      .8078049    .8631699
   _subpop_3 |   .3116783   .0049623      .3019634    .3215617
-------------+------------------------------------------------
_prop_2      |
   _subpop_1 |   .4465001   .0149794      .4171403    .4762378
   _subpop_2 |   .1599003   .0138978        .13432    .1892871
   _subpop_3 |   .6816541   .0049686      .6717621    .6913849
-------------+------------------------------------------------
_prop_3      |
   _subpop_1 |   .0079644   .0040704       .002894    .0217247
   _subpop_2 |   .0027225   .0019217      .0006742    .0109255
   _subpop_3 |   .0066676   .0006743      .0054598    .0081404
--------------------------------------------------------------
```

F.3 Victimization Rates

Victimization rates are calculated from the modified person-level file (for personal crimes) or the modified household-level file (for property crimes) using the Stata *means* command. Since the victimization rate is a function of a constant and two variables, the analysis variable is created as a recode prior to creating the estimates. The victimization count for the victimization type of interest is multiplied by the victimization adjustment factor (*ADJINC_WT*), and this product is multiplied by 1,000. This new variable is used as the analysis variable, as demonstrated in the examples below. Exclusions based on the incident characteristics must be made when calculating victimization summaries, as outlined in ***Section 3.2***. For example, the modified person- and household-level files exclude victimizations occurring outside of the United States. Because there are no dummy records on the modified person and household files, no exclusions are needed to remove dummy records from the analysis.

The modified person and household files contain the victimization counts needed to calculate victimization rates for the most common victimization characteristics analyzed using NCVS data. *Examples 5A-5C* demonstrate the calculation of personal and property victimizations rates that can be calculated directly from the modified person and household level files provided for a single year and *Examples 6A-6B* demonstrate these calculations for pooled year estimates. If an analyst wants to calculate a victimization rate for an incident-level characteristic that is not included on the modified files, preprocessing steps are needed to calculate victimization summaries from the incident-level file and move these summaries to the person file (for personal crimes) or the household file (for property crimes). ***Section 3.2*** documents these steps. *Examples 7A-7B* demonstrate the calculation of personal and property victimization rates for incident characteristics not included on the modified person and household files.

Example 5: Rate of crime, single year

Example 5A is an estimate of an overall personal victimization rate. *Example 5B* is an estimate of a property victimization rate with the inclusion of household-level domain characteristics. *Example 5C* is an estimate of a personal victimization rate, where the domain characteristics of interest are incident-level characteristics. All three estimates are computed with variables available on the modified person and household files.

Example 5A - Rate of simple assaults, 2011

Year(s): 2011
Crime Type(s): simple assault (*sast*)
Domain(s): n/a
Weight: *wgtpercy*
Subpopulation: 2011 (*year*=**2011**)
Calculated Directly from Adjusted Files?: yes

Code:

```
use "<long path of person-level .dta file>" //a
egen strat=group(yr_grp v2117) //b
svyset v2118 [pweight=newwgt], strata(strat) vce(linearized) //c

generate sub=(year==2011) //d
generate vrsast=adjinc_wt*sast*1000 //e
svy, subpop(sub): mean vrsast //f
```

Code Comment(s):

a) Read in input data
b) Create nested stratification recode
c) Specify the NCVS design
d) Create subpopulation indicator
e) Create analysis recode
f) Estimate rate

Output:

```
Survey: Mean estimation

Number of strata =     160      Number of obs    =      814680
Number of PSUs   =     320      Population size  =  1270197175
                                Subpop. no. obs  =      143122
                                Subpop. size     = 257542238.4
                                Design df        =         160

--------------------------------------------------------------
             |             Linearized
             |      Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
      vrsast |   15.36713   1.176853      13.04296     17.6913
--------------------------------------------------------------
```

Example 5B - Rate of household burglary by MSA status, household income, and family structure, 2011

Year(s): 2011
Crime Type(s): household burglary (*hburg*)
Domain(s): MSA status (*msa*); household income (*hincome*); family structure (*fam_structure2*)
Weight: *wgthhcy*
Subpopulation: 2011 (*year*=**2011**)
Calculated Directly from Adjusted Files?: yes

Code:

```
clear
use "<long path of household-level .dta file>"
egen strat=group(yr_grp v2117)
svyset v2118 [pweight=wgthhcy], strata(strat) vce(linearized)

generate sub=(year==2011)
generate vrburg=adjinc_wt*hburg*1000
svy, subpop(sub): mean vrburg, over(msa)
svy, subpop(sub): mean vrburg, over(hincome)
svy, subpop(sub): mean vrburg, over(fam_structure2)
```

Output:

```
Survey: Mean estimation

Number of strata =     160       Number of obs    =     527673
Number of PSUs   =     320       Population size  =  608895975
                                 Subpop. no. obs  =      79802
                                 Subpop. size     =  123038566
                                 Design df        =        160

       Urban: msa = Urban
    Suburban: msa = Suburban
       Rural: msa = Rural

--------------------------------------------------------------
             |             Linearized
        Over |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
vrburg       |
       Urban |   33.48701   2.515425      28.51929    38.45473
    Suburban |   25.44535   2.008611      21.47854    29.41215
       Rural |   33.01906   4.534446      24.06397    41.97414
--------------------------------------------------------------
```

```
. svy, subpop(sub): mean vrburg, over(hincome)
(running mean on estimation sample)

Survey: Mean estimation

Number of strata =     160       Number of obs    =     527673
```

```
Number of PSUs   =     320       Population size  =  608895975
                                 Subpop. no. obs  =      79802
                                 Subpop. size     =  123038566
                                 Design df        =        160

    _subpop_1: hincome = Less than $7,500
    _subpop_2: hincome = $7,500 to $14,999
    _subpop_3: hincome = $15,000 to $24,999
    _subpop_4: hincome = $25,000 to $34,999
    _subpop_5: hincome = $35,000 to $49,999
    _subpop_6: hincome = $50,000 to $74,999
    _subpop_7: hincome = $75,000 or more
    _subpop_8: hincome = Unknown

--------------------------------------------------------------
             |             Linearized
        Over |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
vrburg       |
   _subpop_1 |   67.16497   14.00675      39.50302    94.82692
   _subpop_2 |    58.8451   9.944788      39.20512    78.48508
   _subpop_3 |   42.77625   5.556782      31.80215    53.75034
   _subpop_4 |   33.34014   5.830933      21.82462    44.85566
   _subpop_5 |   26.30117   2.682444       21.0036    31.59873
   _subpop_6 |   21.50634   3.006216      15.56936    27.44332
   _subpop_7 |    12.2887   1.235878      9.847966    14.72944
   _subpop_8 |   29.54958   2.702723      24.21196    34.88719
--------------------------------------------------------------
```

```
. svy, subpop(sub): mean vrburg, over(fam_structure2)
(running mean on estimation sample)

Survey: Mean estimation

Number of strata =     160       Number of obs    =     501100
Number of PSUs   =     320       Population size  =  608892986
                                 Subpop. no. obs  =      79800
                                 Subpop. size     =  123035576
                                 Design df        =        160

    _subpop_1: fam_structure2 = Two or more adults - W/O Childre
    _subpop_2: fam_structure2 = Two or more adults - With Childr
    _subpop_3: fam_structure2 = One Male Adult - W/O Children
    _subpop_4: fam_structure2 = One Male Adult - With Children
    _subpop_5: fam_structure2 = One Female Adult - W/O Children
    _subpop_6: fam_structure2 = One Female Adult - With Children
```

Over	Mean	Linearized Std. Err.	[95% Conf.	Interval]
vrburg				
_subpop_1	24.20889	2.304288	19.65814	28.75963
_subpop_2	29.61002	2.381918	24.90597	34.31408
_subpop_3	30.99701	2.922347	25.22566	36.76836
_subpop_4	46.66971	14.31937	18.39038	74.94905
_subpop_5	30.10567	4.79787	20.63035	39.58099
_subpop_6	61.96275	8.315094	45.54126	78.38424

Example 5C - Rate of violent crime by weapon involvement, injury, and victim-offender relationship, 2011

Year(s): 2011
Crime Type(s): violent victimizations (*rsa+rob+aast+sast*)
Domain(s): weapon involvement (*wpnuse*); injury (*inj*), victim-offender relationship (*rel*)
Weight: *wgtpercy*
Subpopulation: 2011 (*year*=**2011**)
Calculated Directly from Adjusted Files?: yes

Code:

```
clear
use "<long path of person-level .dta file>"
egen strat=group(yr_grp v2117)
svyset v2118 [pweight=wgtpercy], strata(strat) vce(linearized)

generate sub=(year==2011)

generate viol_weap1=(rsa_wpnuse1+rob_wpnuse1+aast_wpnuse1+sast_wpnuse1)*adjinc_wt*1000
//a
generate viol_weap
2=(rsa_wpnuse2+rob_wpnuse2+aast_wpnuse2+sast_wpnuse2)*adjinc_wt*1000 //b
generate viol_weap
3=(rsa_wpnuse3+rob_wpnuse3+aast_wpnuse3+sast_wpnuse3)*adjinc_wt*1000 //c

generate viol_inj1=(rsa_inj1+rob_inj1+aast_inj1+sast_inj1)*adjinc_wt*1000 //d
generate viol_inj2=(rsa_inj2+rob_inj2+aast_inj2+sast_inj2)*adjinc_wt*1000 //e
generate viol_inj3=(rsa_inj3+rob_inj3+aast_inj3+sast_inj3)*adjinc_wt*1000 //f
generate viol_inj4=(rsa_inj4+rob_inj4+aast_inj4+sast_inj4)*adjinc_wt*1000 //g

generate viol_rel1=(rsa_rel1+rob_rel1+aast_rel1+sast_rel1)*adjinc_wt*1000 //h
generate viol_rel2=(rsa_rel2+rob_rel2+aast_rel2+sast_rel2)*adjinc_wt*1000 //i
generate viol_rel3=(rsa_rel3+rob_rel3+aast_rel3+sast_rel3)*adjinc_wt*1000 //j
generate viol_rel4=(rsa_rel4+rob_rel4+aast_rel4+sast_rel4)*adjinc_wt*1000 //k
generate viol_rel5=(rsa_rel5+rob_rel5+aast_rel5+sast_rel5)*adjinc_wt*1000 //l
generate viol_rel6=(rsa_rel6+rob_rel6+aast_rel6+sast_rel6)*adjinc_wt*1000 //m

svy, subpop(sub): mean viol_weap1
svy, subpop(sub): mean viol_weap2
svy, subpop(sub): mean viol_weap3

svy, subpop(sub): mean viol_inj1
svy, subpop(sub): mean viol_inj2
svy, subpop(sub): mean viol_inj3
svy, subpop(sub): mean viol_inj4

svy, subpop(sub): mean viol_rel1
svy, subpop(sub): mean viol_rel2
svy, subpop(sub): mean viol_rel3
svy, subpop(sub): mean viol_rel4
svy, subpop(sub): mean viol_rel5
svy, subpop(sub): mean viol_rel6
```

Code Comment(s):

a) Yes, offender had a weapon
b) No, offender did not have a weapon
c) Do not know if offender had a weapon
d) No injury
e) Serious injury
f) Minor injury
g) Rape without other injuries
h) Intimates
i) Other relatives
j) well known/casual acquaintances
k) Strangers
l) Do not know relationship
m) Do not know number of offenders

Output:

```
Survey: Mean estimation

Number of strata =     160      Number of obs    =      814680
Number of PSUs   =     320      Population size  =  1270197175
                                Subpop. no. obs  =      143122
                                Subpop. size     = 257542238.4
                                Design df        =         160

--------------------------------------------------------------
             |             Linearized
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
viol_weap1   |   4.637771   .3481812      3.950147    5.325395
--------------------------------------------------------------


. svy, subpop(sub): mean viol_weap2
(running mean on estimation sample)

Survey: Mean estimation

Number of strata =     160      Number of obs    =      814680
Number of PSUs   =     320      Population size  =  1270197175
                                Subpop. no. obs  =      143122
                                Subpop. size     = 257542238.4
                                Design df        =         160

--------------------------------------------------------------
             |             Linearized
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
  viol_weap2 |   16.45241   1.176199      14.12953    18.77529
--------------------------------------------------------------




. svy, subpop(sub): mean viol_weap3
(running mean on estimation sample)

Survey: Mean estimation

Number of strata =     160      Number of obs    =      814680
Number of PSUs   =     320      Population size  =  1270197175
                                Subpop. no. obs  =      143122
                                Subpop. size     = 257542238.4
```

```
                                Design df        =         160

------------------------------------------------------------------
             |             Linearized
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------
  viol_weap3 |   1.479021    .300596      .8853732    2.072668
------------------------------------------------------------------


. svy, subpop(sub): mean viol_inj1
(running mean on estimation sample)

Survey: Mean estimation

Number of strata =     160      Number of obs    =      814680
Number of PSUs   =     320      Population size  =  1270197175
                                Subpop. no. obs  =      143122
                                Subpop. size     = 257542238.4
                                Design df        =         160

------------------------------------------------------------------
             |             Linearized
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------
   viol_inj1 |   16.93653   1.225512      14.51627     19.3568
------------------------------------------------------------------


. svy, subpop(sub): mean viol_inj2
(running mean on estimation sample)

Survey: Mean estimation

Number of strata =     160      Number of obs    =      814680
Number of PSUs   =     320      Population size  =  1270197175
                                Subpop. no. obs  =      143122
                                Subpop. size     = 257542238.4
                                Design df        =         160


------------------------------------------------------------------
             |             Linearized
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------
   viol_inj2 |   1.129908   .1913353      .7520398    1.507777
------------------------------------------------------------------

. svy, subpop(sub): mean viol_inj3
(running mean on estimation sample)

Survey: Mean estimation

Number of strata =     160      Number of obs    =      814680
Number of PSUs   =     320      Population size  =  1270197175
                                Subpop. no. obs  =      143122
                                Subpop. size     = 257542238.4
                                Design df        =         160

------------------------------------------------------------------
             |             Linearized
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------
   viol_inj3 |   4.085352    .384823      3.325364    4.845339
------------------------------------------------------------------


. svy, subpop(sub): mean viol_inj4
(running mean on estimation sample)
```

```
Survey: Mean estimation

Number of strata =     160      Number of obs    =      814680
Number of PSUs   =     320      Population size  =  1270197175
                                Subpop. no. obs  =      143122
                                Subpop. size     = 257542238.4
                                Design df        =         160

--------------------------------------------------------------
             |             Linearized
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
   viol_inj4 |   .4174086   .0970625      .2257197    .6090974
--------------------------------------------------------------


. svy, subpop(sub): mean viol_rel1
(running mean on estimation sample)

Survey: Mean estimation

Number of strata =     160      Number of obs    =      814680
Number of PSUs   =     320      Population size  =  1270197175
                                Subpop. no. obs  =      143122
                                Subpop. size     = 257542238.4
                                Design df        =         160

--------------------------------------------------------------
             |             Linearized
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
   viol_rel1 |   3.303427   .4996533      2.316661    4.290193
--------------------------------------------------------------


. svy, subpop(sub): mean viol_rel2
(running mean on estimation sample)

Survey: Mean estimation

Number of strata =     160      Number of obs    =      814680
Number of PSUs   =     320      Population size  =  1270197175
                                Subpop. no. obs  =      143122
                                Subpop. size     = 257542238.4
                                Design df        =         160

--------------------------------------------------------------
             |             Linearized
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
   viol_rel2 |   1.957484   .3548145       1.25676    2.658207
--------------------------------------------------------------


. svy, subpop(sub): mean viol_rel3
(running mean on estimation sample)

Survey: Mean estimation

Number of strata =     160      Number of obs    =      814680
Number of PSUs   =     320      Population size  =  1270197175
                                Subpop. no. obs  =      143122
                                Subpop. size     = 257542238.4
                                Design df        =         160

--------------------------------------------------------------
             |             Linearized
             |       Mean   Std. Err.     [95% Conf. Interval]
```

```
-------------+----------------------------------------------
   viol_rel3 |   6.910004   .7779273      5.373674    8.446334
-------------------------------------------------------------

. svy, subpop(sub): mean viol_rel4
(running mean on estimation sample)

Survey: Mean estimation

Number of strata =     160      Number of obs    =      814680
Number of PSUs   =     320      Population size  =  1270197175
                                Subpop. no. obs  =      143122
                                Subpop. size     = 257542238.4
                                Design df        =         160


-------------------------------------------------------------
             |             Linearized
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------
   viol_rel4 |   8.333804   .6867007      6.977638    9.689971
-------------------------------------------------------------


. svy, subpop(sub): mean viol_rel5
(running mean on estimation sample)

Survey: Mean estimation

Number of strata =     160      Number of obs    =      814680
Number of PSUs   =     320      Population size  =  1270197175
                                Subpop. no. obs  =      143122
                                Subpop. size     = 257542238.4
                                Design df        =         160

-------------------------------------------------------------
             |             Linearized
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------
   viol_rel5 |   .9355984   .1256837      .6873855    1.183811
-------------------------------------------------------------


. svy, subpop(sub): mean viol_rel6
(running mean on estimation sample)

Survey: Mean estimation

Number of strata =     160      Number of obs    =      814680
Number of PSUs   =     320      Population size  =  1270197175
                                Subpop. no. obs  =      143122
                                Subpop. size     = 257542238.4
                                Design df        =         160

-------------------------------------------------------------
             |             Linearized
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------
   viol_rel6 |   1.128885   .2152235      .7038395     1.55393
-------------------------------------------------------------
```

Example 6: Rate of crime, aggregated years

Example 6A is an estimate of a personal victimization rate with person-level domain characteristics of interest. *Example 6B* is an estimate of a property victimization rate with the inclusion of a household-level domain characteristic. Both estimates are computed with variables available on the modified person and household files.

Example 6A - Rate of violent crimes reported to police by sex, age category, race/ethnicity, and marital status, 2009–2011

Year(s): 2009-2011
Crime Type(s): violent victimizations reported to police (*rsa_rpt1+rob_rpt1+aast_rpt1+sast_rpt1*)
Domain(s): sex (*sex*); age category (*agecat*); race/ethnicity (*race_eth*); marital status (*marital*)
Weight: *wgtpercy*
Subpopulation: 2009-2011 ((**2009** <= *year*) and (*year* <=**2011**))
Calculated Directly from Adjusted Files?: yes
Code:

```
replace sub=(year>=2009 & year<=2011)
generate viol_rpt1=(rsa_rpt1+rob_rpt1+aast_rpt1+sast_rpt1)*adjinc_wt*1000

svy, subpop(sub): mean viol_rpt1, over(sex)
svy, subpop(sub): mean viol_rpt1, over(agecat)
svy, subpop(sub): mean viol_rpt1, over(race_eth)
svy, subpop(sub): mean viol_rpt1, over(marital)
```

Output:

```
Survey: Mean estimation

Number of strata =     160      Number of obs    =      814680
Number of PSUs   =     320      Population size  =  1270197175
                                Subpop. no. obs  =      427018
                                Subpop. size     =   767609782
                                Design df        =         160

        Male: sex = Male
      Female: sex = Female

--------------------------------------------------------------
             |             Linearized
        Over |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
viol_rpt1    |
        Male |   9.682951   .5353547      8.625678    10.74022
      Female |   10.80133   .7500923      9.319975    12.28269
--------------------------------------------------------------



. svy, subpop(sub): mean viol_rpt1, over(agecat)
(running mean on estimation sample)

Survey: Mean estimation
```

```
Number of strata =     160      Number of obs    =      814680
Number of PSUs   =     320      Population size  =  1270197175
                                Subpop. no. obs  =      427018
                                Subpop. size     =   767609782
                                Design df        =         160

    _subpop_1: agecat = 12 to 14
    _subpop_2: agecat = 15 to 17
    _subpop_3: agecat = 18 to 20
    _subpop_4: agecat = 21 to 24
    _subpop_5: agecat = 25 to 34
    _subpop_6: agecat = 35 to 49
    _subpop_7: agecat = 50 to 64
    _subpop_8: agecat = 65 or older

------------------------------------------------------------------
             |             Linearized
        Over |       Mean   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------
viol_rpt1    |
   _subpop_1 |   10.14255   1.528288       7.12433    13.16077
   _subpop_2 |   13.04128   1.607437      9.866752    16.21581
   _subpop_3 |     22.577   4.639424      13.41459     31.7394
   _subpop_4 |   16.93357   1.895897      13.18936    20.67778
   _subpop_5 |   14.07592   1.173494      11.75839    16.39346
   _subpop_6 |   11.41399   .8939789      9.648466    13.17951
   _subpop_7 |   6.593241   .5823092      5.443238    7.743245
   _subpop_8 |   1.902745   .3719552       1.16817    2.637319
------------------------------------------------------------------
```

```
. svy, subpop(sub): mean viol_rpt1, over(race_eth)
(running mean on estimation sample)

Survey: Mean estimation

Number of strata =     160      Number of obs    =      814680
Number of PSUs   =     320      Population size  =  1270197175
                                Subpop. no. obs  =      427018
                                Subpop. size     =   767609782
                                Design df        =         160

    _subpop_1: race_eth = Non-Hispanic White
    _subpop_2: race_eth = Non-Hispanic Black
    _subpop_3: race_eth = Hispanic
    _subpop_4: race_eth = Non-Hispanic Other
    _subpop_5: race_eth = Non-Hispanic More than One Race
```

```
-------------------------------------------------------------
             |             Linearized
        Over |       Mean   Std. Err.     [95% Conf. Interval]
-------------+-----------------------------------------------
viol_rpt1    |
   _subpop_1 |   9.591058   .5561921      8.492633    10.68948
   _subpop_2 |   16.32449   2.150383       12.0777    20.57129
   _subpop_3 |   9.110811   .9057507      7.322043    10.89958
   _subpop_4 |   6.215577   1.431772      3.387968    9.043186
   _subpop_5 |   20.78554   3.675282      13.52722    28.04386
-------------------------------------------------------------

. svy, subpop(sub): mean viol_rpt1, over(marital)
(running mean on estimation sample)

Survey: Mean estimation

Number of strata =     160      Number of obs    =      809861
Number of PSUs   =     320      Population size  =  1263685751
                                Subpop. no. obs  =      423279
                                Subpop. size     = 761098358.9
                                Design df        =         160

-------------------------------------------------------------
             |             Linearized
         Over |       Mean   Std. Err.     [95% Conf. Interval]
-------------+-----------------------------------------------
viol_rpt1    |
Never Married |   15.24882   1.034491      13.20581    17.29184
      Married |   5.035802   .3831138       4.27919    5.792414
      Widowed |    3.73865   .9161769      1.929291    5.548009
     Divorced |    18.3727   1.868214      14.68316    22.06224
    Separated |   38.60338   5.885568      26.97997     50.2268
 -------------------------------------------------------------
```

Example 6B - Rate of property crimes reported to police by household income, 2009–2011

Year(s): 2009-2011
Crime Type(s): property crimes reported to police (*hburg_rpt1+tft_rpt1+mvtft_rpt1*)
Domain(s): household income (*hincome*)
Weight: *wgthhcy*
Subpopulation: 2009-2011 ((**2009** <= *year*) and (*year* <=**2011**))
Calculated Directly from Adjusted Files?: yes

Code:

```
clear
use "<long path of household-level .dta file>"
egen strat=group(yr_grp v2117)
svyset v2118 [pweight=wgthhcy], strata(strat) vce(linearized)

generate sub=(year>=2009 & year<=2011)
generate prop_rpt1=(hburg_rpt1+tft_rpt1+mvtft_rpt1)*adjinc_wt*1000

svy, subpop(sub): mean prop_rpt1, over(hincome)
```

Output:

```
Survey: Mean estimation

Number of strata =     160       Number of obs    =     527673
Number of PSUs   =     320       Population size  =  608895975
                                 Subpop. no. obs  =     239205
                                 Subpop. size     =  368251383
                                 Design df        =        160

    _subpop_1: hincome = Less than $7,500
    _subpop_2: hincome = $7,500 to $14,999
    _subpop_3: hincome = $15,000 to $24,999
    _subpop_4: hincome = $25,000 to $34,999
    _subpop_5: hincome = $35,000 to $49,999
    _subpop_6: hincome = $50,000 to $74,999
    _subpop_7: hincome = $75,000 or more
    _subpop_8: hincome = Unknown

--------------------------------------------------------------
             |             Linearized
        Over |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
prop_rpt1    |
   _subpop_1 |   64.05274   4.366416       55.4295    72.67598
   _subpop_2 |   60.84082   3.678508      53.57613    68.10551
   _subpop_3 |   54.83199   2.794859      49.31241    60.35156
   _subpop_4 |   54.83207   3.222495      48.46796    61.19618
   _subpop_5 |   55.92131   3.030289      49.93679    61.90583
   _subpop_6 |   49.01113   2.029558      45.00296    53.01931
   _subpop_7 |   49.94612   1.490552      47.00243    52.88981
   _subpop_8 |   43.61699   1.544342      40.56706    46.66691
--------------------------------------------------------------
```

Example 7: Computing victimization rates based on incident characteristics not included on the provided file

The two examples below estimate personal and property victimization rates for incident-level characteristics not contained on the modified person and household files, and thus require the pre-processing steps outlined in ***Section 3.2***. Stata is used to calculate victimization summaries from the incident-level file and merge incident summaries onto the household and person files. Victimization rates are then calculated from the modified person-level file (for personal crimes) or the modified household-level file (for property crimes). *Example 7A* is an estimate of a personal victimization rate and *Example 7B* is an estimate of a property victimization rate. *Example 7A* is for a single year, and *Example 7B* is for aggregated years.

Example 7A - Rate of violent crimes by location of crime, 2011

Year(s): 2011
Crime Type(s): violent victimizations by location of crime (*violent_home*; *violent_other*)
Domain(s): n/a
Weight: *wgtpercy*
Subpopulation: 2011 (*year*=**2011**)
Calculated Directly from Adjusted Files?: no

Code:

```
clear
use "<long path of incident-level .dta file>"

generate violcr=(inlist(newoff,1,2,3,4)) //a

generate place_inc=. //b
replace place_inc=1 if inlist(v4024,1,2,3,4) //c
replace place_inc=2 if inlist(v4024,5,6,7) //d
replace place_inc=3 if inlist(v4024,8,9,10,11) //e
replace place_inc=4 if inlist(v4024,12,13,14,24,25,26,27) //f
replace place_inc=5 if inlist(v4024,15,16,17) //g
replace place_inc=6 if inlist(v4024,18,19) //h
replace place_inc=7 if inlist(v4024,20,21,22) //i
replace place_inc=8 if inlist(v4024,23) //j
replace place_inc=9 if inlist(v4024,.) //k

generate place=2 //l
replace place=1 if inlist(place_inc,1,2) //m

generate violent_home=(vcrime==1 & place==1 & exclude_outus==0 &
dummy==0) //n
generate violent_other=(vcrime==1 & place==2 & exclude_outus==0 &
dummy==0) //o
```

```
collapse (sum) violent_home violent_other [fw=serieswgt], by(yearq
idhh idper) //p

sort yearq idhh idper
save summary, replace

clear
use "<long path of person-level .dta file>"
egen strat=group(yr_grp v2117)
svyset v2118 [pweight=wgtpercy], strata(strat) vce(linearized)

sort yearq idhh idper

merge m:1 yearq idhh idper using summary //q

generate sub=(year==2011)

replace violent_home=0 if violent_home==.
replace violent_other=0 if violent_other==.

generate vrviolrhome =violent_home*adjinc_wt*1000
generate vrviolother=violent_other*adjinc_wt*1000

svy, subpop(sub): mean vrviolrhome
svy, subpop(sub): mean vrviolother
```

Code Comment(s):

a) Violent crime indicator
b) Location of crime
c) Location: respondent’s home
d) Location: near respondent’s home
e) Location: friend’s home
f) Location: commercial
g) Location: parking lot/garage
h) Location: school
i) Location: open area, on street or public transportation
j) Location: other
k) Location: missing
l) Collapsed location of crime: other
m) Collapsed location of crime: in or near respondent’s home
n) Violent crime at home indicator
o) Violent crime in other location indicator
p) Crime sums by person
q) Bring crime sums onto person-level file

Output:

```
Survey: Mean estimation

Number of strata =     160      Number of obs    =      814680
Number of PSUs   =     320      Population size  =  1270197175
                                Subpop. no. obs  =      143122
                                Subpop. size     = 257542238.4
                                Design df        =         160

--------------------------------------------------------------
             |             Linearized
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
  vrviolrhome|   8.334482   .6313748      7.087579    9.581385
--------------------------------------------------------------


. svy, subpop(sub): mean vrviolother
(running mean on estimation sample)

Survey: Mean estimation


Number of strata =     160      Number of obs    =      814680
Number of PSUs   =     320      Population size  =  1270197175
                                Subpop. no. obs  =      143122
                                Subpop. size     = 257542238.4
                                Design df        =         160

--------------------------------------------------------------
             |             Linearized
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
 vrviolother |   14.23472   1.153994       11.9557    16.51374
--------------------------------------------------------------
```

Example 7B - Rate of property crimes by time of day, 2009-2011

Year(s): 2009-2011
Crime Type(s): property crimes by time of day (*property_day*; *property_night*; *property_missing*)
Domain(s): n/a
Weight: *wgthhcy*
Subpopulation: 2009-2011 ((**2009** <= *year*) and (*year* <=**2011**))
Calculated Directly from Adjusted Files?: no

Code:

```
clear
use "<long path of incident-level .dta file>"

generate pcrime=(inlist(newoff,6,7,8))

//a
generate time_day=.
replace time_day=1 if inlist(v4021b,1,2,3,4)
replace time_day=2 if inlist(v4021b,5,6,7,8)
replace time_day=3 if inlist(v4021b,9,98)

//b
generate property_day=(pcrime==1 & time_day==1 & exclude_outus==0 &
dummy==0)
generate property_night=(pcrime==1 & time_day==2 & exclude_outus==0 &
dummy==0)
generate property_missing=(pcrime==1 & time_day==3 & exclude_outus==0
& dummy==0)

//c
collapse (sum) property_day property_night property_missing
[fw=serieswgt], by(yearq idhh)

sort yearq idhh
save summary, replace

clear
use "<long path of household-level .dta file>"
egen strat=group(yr_grp v2117)
svyset v2118 [pweight=wgthhcy], strata(strat) vce(linearized)

sort yearq idhh

//d
merge m:1 yearq idhh using summary

generate sub=(year>=2009 & year<=2011)

replace property_day=0 if property_day==.
replace property_night=0 if property_night==.
```

```
replace property_missing=0 if property_missing==.

generate vrpropday=property_day*adjinc_wt*1000
generate vrpropnight=property_night*adjinc_wt*1000
generate vrpropmiss=property_missing*adjinc_wt*1000

svy, subpop(sub): mean vrpropday
svy, subpop(sub): mean vrpropnight
svy, subpop(sub): mean vrpropmiss
```

Code Comments:

a) Time of incident variable
b) Property crime indicators by time of incident
c) Crime totals by household
d) Put crime totals on household-level file

Output:

```
Survey: Mean estimation

Number of strata =     160       Number of obs    =     527673
Number of PSUs   =     320       Population size  =  608895975
                                 Subpop. no. obs  =     239205
                                 Subpop. size     =  368251383
                                 Design df        =        160

--------------------------------------------------------------
             |             Linearized
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
   vrpropday |    51.6205    1.22708      49.19713    54.04386
--------------------------------------------------------------

. svy, subpop(sub): mean vrpropnight
(running mean on estimation sample)

Survey: Mean estimation

Number of strata =     160       Number of obs    =     527673
Number of PSUs   =     320       Population size  =  608895975
                                 Subpop. no. obs  =     239205
                                 Subpop. size     =  368251383
                                 Design df        =        160

--------------------------------------------------------------
             |             Linearized
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
 vrpropnight |   56.47746   1.331527      53.84782    59.10709
--------------------------------------------------------------




. svy, subpop(sub): mean vrpropmiss
(running mean on estimation sample)

Survey: Mean estimation

Number of strata =     160       Number of obs    =     527673
Number of PSUs   =     320       Population size  =  608895975
                                 Subpop. no. obs  =     239205
```

```
                                Subpop. size     =  368251383
                                Design df        =        160

------------------------------------------------------------
             |             Linearized
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------
  vrpropmiss |   24.14251   .8068977      22.54897    25.73606
------------------------------------------------------------
```

F.4 Identifying Low Quality Estimates

This section demonstrates how to implement the recommendations for identifying low-quality estimates and the rounding rules discussed in ***Section 4*** of the user's guide. Three types of data are needed to identify and flag low quality estimates, regardless of the estimate type: the estimate, the standard error of the estimate, and the unweighted sample size of the estimate. Because *Examples 1-7* provided details for obtaining estimates and standard errors, the examples in this section will focus primarily on calculating the percent relative standard error (RSE), calculating unweighted sample sizes, identifying estimates that should be flagged as unreliable, and verifying that estimates rounding to zero are not presented.

Unweighted sample sizes should take into account the series adjustment. For example, a series victimization with a series count of seven would count as seven victimizations in the unweighted sample size, while a non-series victimization would only count as one. This series adjustment is already incorporated in the victimization counts on the modified person and household-level files, so sample sizes for victimization rates are obtained by taking unweighted sums of victimization counts with the specified characteristic(s) of interest. However, for victimization totals and proportions, the series count must be included in the calculation of the sample size, as demonstrated in the examples below. Each example specifies the number of decimals to be displayed in the resulting report or presentation so that rounding rules can be implemented. *Example 8A* demonstrates implementation of flagging and rounding rules for victimization totals, *Example 8B* for victimization proportions, and *Example 8C* for victimization rates.

Example 8A - Total number of personal thefts by sex and race/ethnicity, 2011

Year(s): 2011
Crime Type: personal thefts (*ptft*)
Domain(s): sex (*sex*) and race/ethnicity (*race_eth*)
Weight: *newwgt*
Subpopulation: within the United States (*exclude_outUS*=**0**); non-dummy records (*dummy*=**0**); 2011 (*year*=**2011**)

Code:

```
use "<long path of incident-level .dta file>"
egen strat=group(yr_grp v2117)
svyset v2118 [pweight=newwgt], strata(strat) vce(linearized)

generate sub=(exclude_outus==0 & dummy==0 & year==2011)
svy, subpop(sub): total ptft, over(sex)
svy, subpop(sub): total ptft, over(race_eth)

svyset v2118 [pweight=serieswgt], strata(strat) vce(linearized) //a

svy, subpop(sub): total ptft, over(sex)
svy, subpop(sub): total ptft, over(race_eth)
```

Code Comment(s):

a) Weight by the series weight to obtain the series-adjusted count of victimizations

Output:

```
. svy, subpop(sub): total ptft, over(sex)
(running total on estimation sample)

Survey: Total estimation

Number of strata =     160       Number of obs    =      46677
Number of PSUs   =     320       Population size  =  144563822
                                 Subpop. no. obs  =       7255
                                 Subpop. size     = 23041440.8
                                 Design df        =        160

         Male: sex = Male
       Female: sex = Female

--------------------------------------------------------------
             |             Linearized
        Over |      Total   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
ptft         |
        Male |   53583.48    18913.9      16230.39    90936.56
      Female |   112186.7   19313.73      74043.95    150329.4
--------------------------------------------------------------
Note: 144 strata omitted because they contain no subpopulation
      members.
```

```
. svy, subpop(sub): total ptft, over(race_eth)
(running total on estimation sample)

Survey: Total estimation

Number of strata =     160       Number of obs    =      46677
Number of PSUs   =     320       Population size  =  144563822
```

```
                                Subpop. no. obs  =       7255
                                Subpop. size     = 23041440.8
                                Design df        =        160

    _subpop_1: race_eth = Non-Hispanic White
    _subpop_2: race_eth = Non-Hispanic Black
    _subpop_3: race_eth = Hispanic
    _subpop_4: race_eth = Non-Hispanic Other
    _subpop_5: race_eth = Non-Hispanic More than One Race
------------------------------------------------------------------------------
             |             Linearized
        Over |      Total   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ptft         |
   _subpop_1 |   83955.56   21478.97      41536.71    126374.4
   _subpop_2 |   29506.98   10989.44      7803.916    51210.05
   _subpop_3 |   30286.66    10488.6      9572.709    51000.61
   _subpop_4 |   13519.07   5943.552      1781.144       25257
   _subpop_5 |   8501.872   5281.778      -1929.12    18932.86
------------------------------------------------------------------------------
Note: 144 strata omitted because they contain no subpopulation
      members.


. svy, subpop(sub): total ptft, over(sex)
(running total on estimation sample)

Survey: Total estimation

Number of strata =     160          Number of obs    =   46677
Number of PSUs   =     320          Population size  =   60616
                                    Subpop. no. obs  =    7255
                                    Subpop. size     =    8064
                                    Design df        =     160

         Male: sex = Male
       Female: sex = Female

------------------------------------------------------------------------------
             |             Linearized
        Over |      Total   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ptft         |
        Male |         16   4.898979      6.324998      25.675
      Female |         35   5.567764      24.00421    45.99579
------------------------------------------------------------------------------
Note: 144 strata omitted because they contain no subpopulation
      members.


. svy, subpop(sub): total ptft, over(race_eth)
(running total on estimation sample)

Survey: Total estimation

Number of strata =     160          Number of obs    =   46677
Number of PSUs   =     320          Population size  =   60616
                                    Subpop. no. obs  =    7255
                                    Subpop. size     =    8064
                                    Design df        =     160
    _subpop_1: race_eth = Non-Hispanic White
    _subpop_2: race_eth = Non-Hispanic Black
    _subpop_3: race_eth = Hispanic
    _subpop_4: race_eth = Non-Hispanic Other
    _subpop_5: race_eth = Non-Hispanic More than One Race


------------------------------------------------------------------------------
             |             Linearized
```

```
        Over |      Total   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
ptft         |
   _subpop_1 |         24    5.09902      13.92994    34.07006
   _subpop_2 |          9          3      3.075295     14.9247
   _subpop_3 |          9   2.645751      3.774902     14.2251
   _subpop_4 |          6    2.44949      1.162499     10.8375
   _subpop_5 |          3   1.732051     -.4206298     6.42063
--------------------------------------------------------------
Note: 144 strata omitted because they contain no subpopulation
      members.
```

Identifying Unreliable Estimates:

As outlined in ***Section 4***, it is recommended that estimated totals meeting either of the following criteria be identified as unreliable. In addition, any estimate rounding to zero should not be displayed.

- RSE > 30%
- Count sample size ≤ 10

The percent relative standard error of an estimated total is calculated as:

$$\%\, RSE(Total) = \frac{SE(Total)}{Total} * 100$$

Estimated totals, standard errors, and unweighted sample sizes from the output above are included in ***Table F-1*** (in the "Total," "SE(Total)," and "n" columns, respectively). Percent RSEs are calculated based on the formula above. Based on the recommendations, five of the eight estimates are flagged as unreliable. The estimate for males is flagged because the percent RSE is greater than 30 percent, while the other four estimates are flagged both because their RSEs exceed 30 percent and because their unweighted sample sizes are less than or equal to 10. All estimates can be displayed because the analyst plans to display estimates to the nearest whole number, and no estimates round to zero.

Table F-1. Identifying Low Quality Victimization Totals

Domain	Total	SE(Total)	n	%RSE (Total)	Flag as Unreliable?	Rounds to Zero?
Overall	165770.15	24276.55	51	14.64		
Male	53583.48	18913.90	16	35.30	√	
Female	112186.67	19313.73	35	17.22		
Non-Hispanic White	83955.56	21478.97	24	25.58		
Non-Hispanic Black	29506.98	10989.44	9	37.24	√	
Hispanic	30286.66	10488.60	9	34.63	√	
Non-Hispanic Other	13519.07	5943.55	6	43.96	√	
Non-Hispanic More than One Race	8501.87	5281.78	3	62.12	√	

Example 8B – Distribution of motor vehicle theft in the Northeast Census Region by report status, 2011

Year(s): 2011
Domain(s): Northeast Census Region (*region*=1); motor vehicle theft (*mvtft*=1)
Covariate(s): report status (*notify*)
Weight: *newwgt*
Subpopulation: within the United States (*exclude_outUS*=**0**); non-dummy records (*dummy*=**0**); 2011 (*year*=**2011**); Northeast Census Region (*region*=1); motor vehicle theft (*mvtft*=1)

Code:

```
replace sub=(exclude_outus==0 & dummy==0 & year==2011 & ///
      region==1 & mvtft==1)

svy, subpop(sub): prop notify

tabulate notify if sub==1 [fweight=serieswgt] //a
```

Code Comment(s):

a) Weight by the series weight to obtain the series-adjusted count of victimizations. Because flagging rules are based on the denominator sample size, only the overall sample size is needed (not estimates by report status).

Output:

```
. svy, subpop(sub): prop notify
(running proportion on estimation sample)

Survey: Proportion estimation

Number of strata =      12         Number of obs    =     2590
Number of PSUs   =      24         Population size  =  8426594
                                   Subpop. no. obs  =       19
                                   Subpop. size     = 57944.13
                                   Design df        =       12

      _prop_1: notify = reported to police
      _prop_2: notify = not reported to police
      _prop_3: notify = do not know

--------------------------------------------------------------
             |             Linearized
             | Proportion   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
notify       |
     _prop_1 |   .9260112   .0540324      .6918327    .9858703
     _prop_2 |   .0739888   .0540324      .0141297    .3081673
     _prop_3 |          .  (no observations)
--------------------------------------------------------------
Note: 292 strata omitted because they contain no subpopulation
      members.

. tabulate notify if sub==1 [fweight=serieswgt]

               notify |      Freq.     Percent        Cum.
----------------------+-----------------------------------
    reported to police |         17       89.47       89.47
```

```
not reported to police |          2       10.53      100.00
-----------------------+-----------------------------------
                 Total |         19      100.00
```

Identifying Unreliable Estimates:

As outlined in ***Section 4***, it is recommended that victimization percentages (P) meeting either of the following criteria be flagged as unreliable. In addition, any estimate rounding to zero should not be displayed.

1. RSE > 30%, where
    a. If the percentage is ≤ 50%, use the RSE of log(P)
    b. If the percentage is > 50%, use the RSE of log(100-P); or
2. Denominator sample size ≤ 10

The percent relative standard errors of the log of the percentages (P) and (1-P) are calculated as:

$$\%\, RSE(\log(P)) = \frac{SE(P)}{P * abs(\log\left(\frac{P}{100}\right))} * 100$$

$$\%\, RSE(\log(100-P)) = \frac{SE(P)}{(100-P) * abs(\log\left(1-\frac{P}{100}\right))} * 100$$

Estimated percentages, standard errors, and the unweighted denominator sample size from the output above are included in ***Table F-2*** (in the "P," "SE(P)," and "n (denom)" columns, respectively). Percent RSEs are calculated based on the formulas above. Based on these recommendations, none of the estimates are flagged as unreliable. The denominator sample size is 19, which is greater than the recommended threshold of 10. The percent RSEs for the reported and not reported estimates are 28.03, which are below the 30 percent threshold. However, because no respondents reported that they did not know whether or not the motor vehicle theft was reported to the police, the estimated percent is 0. This estimate should not be presented because it rounds to zero.

Table F-2. Identifying Low Quality Victimization Percentages

Estimate	P	SE(P)	n (denom)	%RSE(P) OR %RSE(100-P)	Flag as Unreliable?	Rounds to Zero?
Reported to Police	92.60	5.40	19	28.03		
Not Reported to Police	7.40	5.40	19	28.03		
Do Not Know	0.00	-	19	-		√

Example 8C – Rate of rape/sexual assault by sex, 2011

Year(s): 2011
Crime Type(s): rape/sexual assault (*rsa*)
Domain(s): sex (*sex*)
Weight: *wgtpercy*
Subpopulation: 2011 (*year*=**2011**)
Calculated Directly from Adjusted Files?: yes
Code:

```
clear
use "<long path of person-level .dta file>"
egen strat=group(yr_grp v2117)
svyset v2118 [pweight=wgtpercy], strata(strat) vce(linearized)

generate sub=(year==2011)
generate vrrsa=adjinc_wt*rsa*1000

svy, subpop(sub): mean vrrsa
svy, subpop(sub): mean vrrsa, over(sex)

svyset v2118, strata(strat) vce(linearized) //a

svy, subpop(sub): total rsa, over(sex)
```

Code Comment(s):

a) Unweighted because the victimization count RSA already includes the series adjustment

Output:

```
. svy, subpop(sub): mean vrrsa
(running mean on estimation sample)

Survey: Mean estimation

Number of strata =     160      Number of obs    =      814680
Number of PSUs   =     320      Population size  =  1270197175
                                Subpop. no. obs  =      143122
                                Subpop. size     = 257542238.4
                                Design df        =         160

--------------------------------------------------------------
             |             Linearized
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
       vrrsa|    .948146   .1625289      .6271673    1.269125
--------------------------------------------------------------




. svy, subpop(sub): mean vrrsa, over(sex)
(running mean on estimation sample)
Survey: Mean estimation

Number of strata =     160      Number of obs    =      814680
```

```
Number of PSUs   =     320      Population size  =  1270197175
                                Subpop. no. obs  =      143122
                                Subpop. size     = 257542238.4
                                Design df        =         160

           1: sex = 1
           2: sex = 2

------------------------------------------------------------------
             |             Linearized
        Over |       Mean   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------
  vrrsa      |
           1 |   .2760193   .0897017      .0988673    .4531713
           2 |   1.592869   .3323435      .9365229    2.249214
------------------------------------------------------------------


. svy, subpop(sub): total rsa, over(sex)
(running total on estimation sample)

Survey: Total estimation

Number of strata =     160          Number of obs    =  814680
Number of PSUs   =     320          Population size  =  814680
                                    Subpop. no. obs  =  162867
                                    Subpop. size     =  162867
                                    Design df        =     160

           1: sex = 1
           2: sex = 2

------------------------------------------------------------------
             |             Linearized
        Over |      Total   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------
rsa          |
           1 |         10   2.828427      4.414135    15.58587
           2 |         69   15.32971      38.72533    99.27467
------------------------------------------------------------------
```

Identifying Unreliable Estimates:

As outlined in ***Section 4***, it is recommended that victimization rates meeting either of the following criteria be flagged as unreliable. In addition, any estimate rounding to zero should not be displayed.

- RSE > 30%
- Numerator sample size ≤ 10

The percent relative standard errors of the victimization rates (VR) are calculated as:

$$\%\, RSE(VR) = \frac{SE(VR)}{VR} * 100$$

Estimates, standard errors, and unweighted sample sizes from the output above are included in ***Table F-3*** (in the "VR," "SE(VR)," and "n (num)" columns, respectively). Percent RSEs are calculated based on the formula above. Based on these recommendations, the estimated victimization rate for males is flagged as unreliable because the percent RSE is greater than 30 percent and the sample size of the numerator is less than or equal to 10. Because neither estimate would round to zero when one decimal place is displayed, all three estimates can be presented.

Table F-3. Identifying Low Quality Victimization Rates

Domain	VR	SE(VR)	n (num)	%RSE(VR)	Flag as Unreliable?	Rounds to Zero?
Overall	0.94815	0.16253	79	17.14		
Male	0.27602	0.0897	10	32.50	√	
Female	1.59287	0.33234	69	20.86		