Analysis Example-Stata 11.0 sgmediation Command with Survey Data Correction
March 25, 2011

This example demonstrates the use of the Stata 11.1 sgmediation command with survey correction and a subpopulation indicator.

The NCS-R data is used in this example. The model is defined as the dependent variable=household income, independent variable=age and mediation variable=obesity status in 6 categories. This model can be regarded as age (IV) influences obesity (MV) which in turn influences household income (DV).

This command performs the complex sample adjusted Sobel-Goodman test to test whether a mediator carries the influence of an independent variable to a dependent variable. See the Stata help for sgmediation for details on this approach. The use of a subpop statement is also included in this example. Use of the svyset command must precede the modeling.

```
. use f:\brahms\applied_analysis_book\ncsrsubset_dec15.dta, clear
. * set survey variables
. svyset seclustr [pweight=ncsrwtlg], strata(sestrat) vce(linearized) singleunit(missing)
      pweight: ncsrwtlg
          VCE: linearized
  Single unit: missing
     Strata 1: sestrat
         SU 1: seclustr
        FPC 1: <zero>
. svydes sestrat seclustr
Survey: Describing stage 1 sampling units
      pweight: ncsrwtlg
          VCE: linearized
  Single unit: missing
     Strata 1: sestrat
         SU 1: seclustr
        FPC 1: <zero>
```

Stratum	#Units included	#Units omitted	#Obs with complete data	#Obs with missing data	#Obs per included Unit: min	mean	max
1	2	0	44	0	22	22.0	22
2	2	0	53	0	26	26.5	27
3	2	0	77	0	31	38.5	46
4	2	0	87	0	42	43.5	45
5	2	0	72	0	32	36.0	40
6	2	0	68	0	31	34.0	37
7	2	0	127	0	61	63.5	66
8	2	0	84	0	38	42.0	46
9	2	0	89	0	44	44.5	45
10	2	0	78	0	23	39.0	55
11	2	0	57	0	28	28.5	29
12	2	0	61	0	24	30.5	37
13	2	0	63	0	26	31.5	37
14	2	0	45	0	20	22.5	25
15	2	0	73	0	36	36.5	37
16	2	0	66	0	30	33.0	36
17	2	0	41	0	18	20.5	23
18	2	0	55	0	26	27.5	29
19	2	0	62	0	29	31.0	33
20	2	0	159	0	70	79.5	89
21	2	0	187	0	75	93.5	112
22	2	0	178	0	82	89.0	96
23	2	0	186	0	84	93.0	102
24	2	0	197	0	94	98.5	103
25	2	0	264	0	124	132.0	140
26	2	0	155	0	58	77.5	97
27	2	0	172	0	79	86.0	93
28	2	0	117	0	45	58.5	72
29	2	0	199	0	99	99.5	100
30	2	0	121	0	57	60.5	64
31	2	0	267	0	131	133.5	136
32	2	0	191	0	86	95.5	105
33	2	0	82	0	37	41.0	45
34	2	0	236	0	108	118.0	128
35	2	0	236	0	94	118.0	142

```
      36          2          0        203          0        100     101.5       103
      37          2          0        218          0        106     109.0       112
      38          2          0        197          0         84      98.5       113
      39          2          0        215          0         76     107.5       139
      40          2          0        164          0         66      82.0        98
      41          2          0        211          0         94     105.5       117
      42          2          0        235          0        116     117.5       119
--------  ---------  ---------  ---------  ---------  ---------  ---------  ---------
      42         84          0       5692          0         18      67.8       142
                                 ------------------
                                       5692
                                       3590 = #Obs with missing values in the
                                     --------   survey characteristcs
                                       9282


. * do analysis of household income predicted by age and mediated by obesity status, with survey correction
. sgmediation hhinc, mv(obese6ca) iv(age) prefix(svy:)

Model with dv regressed on iv (path c)
(running regress on estimation sample)

Survey: Linear regression

Number of strata   =        42                  Number of obs      =      5594
Number of PSUs     =        84                  Population size    = 5598.7597
                                                Design df          =        42
                                                F(   1,     42)    =      7.34
                                                Prob > F           =    0.0097
                                                R-squared          =    0.0034

------------------------------------------------------------------------------
             |             Linearized
       hhinc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -157.7628   58.22263    -2.71   0.010    -275.2608   -40.26475
       _cons |   66324.58   3603.259    18.41   0.000     59052.91    73596.26
------------------------------------------------------------------------------

Model with mediator regressed on iv (path a)
(running regress on estimation sample)

Survey: Linear regression

Number of strata   =        42                  Number of obs      =      5594
Number of PSUs     =        84                  Population size    = 5598.7597
                                                Design df          =        42
                                                F(   1,     42)    =     13.10
                                                Prob > F           =    0.0008
                                                R-squared          =    0.0045

------------------------------------------------------------------------------
             |             Linearized
    obese6ca |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0041952   .0011593     3.62   0.001     .0018556    .0065347
       _cons |   2.755127   .0567891    48.52   0.000     2.640522    2.869732
------------------------------------------------------------------------------
```

```
Model with dv regressed on mediator and iv (paths b and c')
(running regress on estimation sample)

Survey: Linear regression

Number of strata   =        42                  Number of obs      =      5594
Number of PSUs     =        84                  Population size    = 5598.7597
                                                Design df          =        42
                                                F(   2,     41)    =      4.75
                                                Prob > F           =    0.0140
                                                R-squared          =    0.0045

------------------------------------------------------------------------------
             |             Linearized
       hhinc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    obese6ca |  -1458.277   663.4159    -2.20   0.034    -2797.105   -119.4494
         age |  -151.6451     57.845    -2.62   0.012     -268.381   -34.90914
       _cons |   70342.32   4533.173    15.52   0.000     61194.01    79490.64
------------------------------------------------------------------------------

Sobel-Goodman Mediation Tests

             Coef         Std Err     Z           P>|Z|
Sobel       -6.1177002    3.2563506  -1.879      .06028571
Goodman-1   -6.1177002    3.3459401  -1.828      .06749029
Goodman-2   -6.1177002    3.1642255  -1.933      .05318747

Indirect effect = -6.1177002
  Direct effect = -151.64508
   Total effect = -157.76278

Proportion of total effect that is mediated:  .03877784
Ratio of indirect to direct effect:      .04034223
```

```
. * repeat analysis with female subpopulation indicator
. sgmediation hhinc, mv(obese6ca) iv(age) prefix(svy, subpop(sexf):)

Model with dv regressed on iv (path c)
(running regress on estimation sample)

Survey: Linear regression

Number of strata   =        42                  Number of obs      =      5594
Number of PSUs     =        84                  Population size    = 5598.7597
                                                Subpop. no. of obs =      3251
                                                Subpop. size       = 2969.2933
                                                Design df          =        42
                                                F(   1,     42)    =     10.65
                                                Prob > F           =    0.0022
                                                R-squared          =    0.0077

------------------------------------------------------------------------------
             |             Linearized
       hhinc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -221.5794   67.89465    -3.26   0.002    -358.5964   -84.56249
       _cons |   62793.69   4159.279    15.10   0.000     54399.93    71187.46
------------------------------------------------------------------------------

Model with mediator regressed on iv (path a)
(running regress on estimation sample)

Survey: Linear regression

Number of strata   =        42                  Number of obs      =      5594
Number of PSUs     =        84                  Population size    = 5598.7597
                                                Subpop. no. of obs =      3251
                                                Subpop. size       = 2969.2933
                                                Design df          =        42
                                                F(   1,     42)    =     16.34
                                                Prob > F           =    0.0002
                                                R-squared          =    0.0083

------------------------------------------------------------------------------
             |             Linearized
    obese6ca |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0060273   .0014912     4.04   0.000      .003018    .0090366
       _cons |   2.596518   .0738942    35.14   0.000     2.447393    2.745642
------------------------------------------------------------------------------

Model with dv regressed on mediator and iv (paths b and c')
(running regress on estimation sample)

Survey: Linear regression

Number of strata   =        42                  Number of obs      =      5594
Number of PSUs     =        84                  Population size    = 5598.7597
                                                Subpop. no. of obs =      3251
                                                Subpop. size       = 2969.2933
                                                Design df          =        42
                                                F(   2,     41)    =     12.08
                                                Prob > F           =    0.0001
                                                R-squared          =    0.0156

------------------------------------------------------------------------------
             |             Linearized
       hhinc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    obese6ca |  -3384.067   795.4815    -4.25   0.000    -4989.413    -1778.72
         age |  -201.1828   67.63443    -2.97   0.005    -337.6746   -64.69096
       _cons |   71580.48   5016.543    14.27   0.000     61456.69    81704.28
------------------------------------------------------------------------------
```

Sobel-Goodman Mediation Tests

	Coef	Std Err	Z	P>|Z|
Sobel	-20.396679	6.9607835	-2.93	.00338714
Goodman-1	-20.396679	7.0611319	-2.889	.0038698
Goodman-2	-20.396679	6.8589672	-2.974	.00294209

Indirect effect = -20.396679
Direct effect = -201.18276
Total effect = -221.57944

Proportion of total effect that is mediated: .09205132
Ratio of indirect to direct effect: .10138383