```
------------------------------------------------------------------------------------------------------------------------------------------------------------------------
      name:  <unnamed>
       log:  D:\Jason\workshop\regression analysis 2022\regression2.log
  log type:  text
 opened on:   3 Oct 2022, 10:52:37

. webuse highschool, clear

. cd "D:\Jason\workshop\regression analysis 2022"
D:\Jason\workshop\regression analysis 2022

. 
. **************************************
. * Continuous dependent variable
. ***************************************
. 
. ******************************************************************************************
. * Questions 1: Determine if X2 is an important predicator when X1 is already in the model
. ******************************************************************************************
. 
. 
.          ****************************************
.          * 1.1 Regression without complex survey data
.          *****************************************
. 
.                  ****************************************************
.                  * 1.1.0 Regression without using the weight variable
.                  ****************************************************
. 
.                  reg weight height i.race

      Source |       SS           df       MS      Number of obs   =     4,071
-------------+----------------------------------   F(3, 4067)      =    432.31
       Model |  1166634.57         3   388878.19   Prob > F        =    0.0000
    Residual |  3658378.15     4,067  899.527452   R-squared       =    0.2418
-------------+----------------------------------   Adj R-squared   =    0.2412
       Total |  4825012.72     4,070  1185.50681   Root MSE        =    29.992

------------------------------------------------------------------------------
      weight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      height |   .6609619   .0188005    35.16   0.000     .6241025    .6978212
             |
        race |
      Black  |   7.697073   1.500755     5.13   0.000     4.754771    10.63938
      Other  |     -12.28   3.654374    -3.36   0.001    -19.44457   -5.115422
             |
       _cons |  -125.0944   8.110502   -15.42   0.000    -140.9954   -109.1934
------------------------------------------------------------------------------

. 
.                  ************************************************************************
.                  * 1.1.1. Regression using a personal weight variable for the whole sample
.                  ************************************************************************
. 
.                  reg weight height i.race [pw=sampwgt]
(sum of wgt is 8,000,000)

Linear regression                               Number of obs     =      4,071
                                                F(3, 4067)        =     376.04
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2839
                                                Root MSE          =     29.051
```

```
------------------------------------------------------------------------------
             |               Robust
      weight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      height |   .7132085    .021832    32.67   0.000     .6704059    .7560112
             |
        race |
      Black  |    6.90636   2.055533     3.36   0.001     2.876389    10.93633
      Other  |  -8.328094   2.618243    -3.18   0.001    -13.46128   -3.194905
             |
       _cons |  -148.7428    9.43488   -15.77   0.000    -167.2404   -130.2453
------------------------------------------------------------------------------

.
.                ************************************************************** *********
.                * 1,1,2. Regression using a personal weight variable and adjusted for
.                * dependent observations from the same schools.
.                ************************************************************************
.
.                 reg weight height i.race [pw=sampwgt], cluster(school)
(sum of wgt is 8,000,000)

Linear regression                               Number of obs     =      4,071
                                                F(3, 9)           =     262.96
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2839
                                                Root MSE          =     29.051

                                (Std. Err. adjusted for 10 clusters in school)
------------------------------------------------------------------------------
             |               Robust
      weight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      height |   .7132085   .0267064    26.71   0.000     .6527946    .7736225
             |
        race |
      Black  |    6.90636   2.001967     3.45   0.007     2.377597    11.43512
      Other  |  -8.328094    2.41195    -3.45   0.007     -13.7843   -2.871885
             |
       _cons |  -148.7428   11.50666   -12.93   0.000    -174.7727    -122.713
------------------------------------------------------------------------------

.
.
.                 ************************************************************************
.                 * 1.1.3. Regression using a personal weight variable and adjusted for
.                 *dependent observations from same schools among female respondents only
.                 ************************************************************************
.
.                 reg weight height i.race if sex ==2 [pw=sampwgt], cluster(school)
(sum of wgt is 4,151,978.6416772)

Linear regression                               Number of obs     =      2,133
                                                F(3, 9)           =      29.63
                                                Prob > F          =     0.0001
                                                R-squared         =     0.0720
                                                Root MSE          =      30.32

                                (Std. Err. adjusted for 10 clusters in school)
------------------------------------------------------------------------------
             |               Robust
      weight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
```

```
-------------+----------------------------------------------------------------
      height |   .4122218   .0613759     6.72   0.000     .2733797    .5510638
             |
        race |
      Black  |   13.97939   3.113373     4.49   0.002     6.936449    21.02233
      Other  |  -6.629714   2.649844    -2.50   0.034    -12.62408   -.6353504
             |
       _cons |  -26.49653   25.33738    -1.05   0.323    -83.81367    30.82062
------------------------------------------------------------------------------

.
.
.          ****************************************
.          * 1.2. Regression with complex survey data
.          *****************************************
.          svydes

Survey: Describing stage 1 sampling units

      pweight: sampwgt
          VCE: linearized
  Single unit: missing
     Strata 1: state
         SU 1: county
        FPC 1: ncounties
     Strata 2: <one>
         SU 2: school
        FPC 2: nschools

                                      #Obs per Unit
                              ----------------------------
Stratum   #Units     #Obs      min       mean      max
--------  --------  --------  --------  --------  --------
       1         2        92        34      46.0        58
       2         2       112        51      56.0        61
       3         2        43        18      21.5        25
       4         2        37        14      18.5        23
       5         2        96        38      48.0        58
       6         2        76        27      38.0        49
       7         2        97        47      48.5        50
       8         2       103        51      51.5        52
       9         2        69        33      34.5        36
      10         2        50        17      25.0        33
      11         2        72        30      36.0        42
      12         2        76        34      38.0        42
      13         2       135        54      67.5        81
      14         2       103        43      51.5        60
      15         2        66        28      33.0        38
      16         2        72        31      36.0        41
      17         2        60        26      30.0        34
      18         2        58        25      29.0        33
      19         2        72        35      36.0        37
      20         2        83        38      41.5        45
      21         2        80        32      40.0        48
      22         2        76        37      38.0        39
      23         2        81        38      40.5        43
      24         2        72        36      36.0        36
      25         2        74        34      37.0        40
      26         2        89        33      44.5        56
      27         2       107        47      53.5        60
      28         2       126        58      63.0        68
      29         2        90        42      45.0        48
      30         2        87        42      43.5        45
```

```
      31         2        84        38      42.0        46
      32         2        95        40      47.5        55
      33         2        97        44      48.5        53
      34         2        82        37      41.0        45
      35         2        98        36      49.0        62
      36         2        95        42      47.5        53
      37         2        53        23      26.5        30
      38         2        75        37      37.5        38
      39         2        61        28      30.5        33
      40         2        54        25      27.0        29
      41         2        73        28      36.5        45
      42         2        79        36      39.5        43
      43         2       108        48      54.0        60
      44         2        76        37      38.0        39
      45         2       107        45      53.5        62
      46         2       115        56      57.5        59
      47         2        67        28      33.5        39
      48         2        56        23      28.0        33
      49         2        78        39      39.0        39
      50         2        64        31      32.0        33
--------  --------  --------  --------  --------  --------
      50       100     4,071        14      40.7        81

.         svyset, clear

.         quietly svyset county [pw = sampwgt], fpc(ncounties) strata(state)|| school, fpc(nschools)

.
.
.                 *********************************************
.                 * 1.2.1. Regression using complex survey data  for the whole sample
.                 *********************************************
.
.                 svy: reg weight height i.race
(running regress on estimation sample)

Survey: Linear regression

Number of strata   =        50                 Number of obs     =      4,071
Number of PSUs     =       100                 Population size   =  8,000,000
                                               Design df         =         50
                                               F(   3,     48)   =     207.69
                                               Prob > F          =     0.0000
                                               R-squared         =     0.2839

------------------------------------------------------------------------------
             |             Linearized
      weight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      height |   .7132085    .029109    24.50   0.000     .6547413    .7716758
             |
        race |
      Black  |    6.90636   2.477742     2.79   0.007     1.929669    11.88305
      Other  |  -8.328094   2.557726    -3.26   0.002    -13.46544   -3.190751
             |
       _cons |  -148.7428    12.3825   -12.01   0.000    -173.6138   -123.8719
------------------------------------------------------------------------------

.
.
.                 ***********************************************************************
.                 * 1,2,2. Regression using complex survey data and adjusted for
.                 * dependent observations from the same schools.
```

```
.                 *************************************************************************
.
.                 * Cluster option is not needed because svy command already take
.                 * into account the dependences among different levels of sampling units
.
.
.                 **************************************************************************
.                 * 1.2.3. Regression using a personal weight variable and adjusted for
.                 *dependent observations from same samping units among female respondents only
.                 **************************************************************************
.
.                 *correct specification*
.
.
.                 svy, subpop(if sex==2): reg weight height i.race
(running regress on estimation sample)

Survey: Linear regression

Number of strata   =        50                  Number of obs     =      4,071
Number of PSUs     =       100                  Population size   =  8,000,000
                                                Subpop. no. obs   =      2,133
                                                Subpop. size      =  4,151,979
                                                Design df         =         50
                                                F(   3,     48)   =      30.05
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0720

------------------------------------------------------------------------------
             |             Linearized
      weight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      height |   .4122218   .0576472     7.15   0.000     .2964339    .5280096
             |
        race |
      Black  |   13.97939   3.538247     3.95   0.000      6.87261    21.08616
      Other  |  -6.629714   2.787416    -2.38   0.021     -12.2284   -1.031025
             |
       _cons |  -26.49653    23.6053    -1.12   0.267    -73.90917    20.91612
------------------------------------------------------------------------------

.
.                 *incorrect specification*
.                 svy: reg weight height i.race if sex==2
(running regress on estimation sample)

Survey: Linear regression

Number of strata   =        50                  Number of obs     =      2,133
Number of PSUs     =       100                  Population size   =  4,151,979
                                                Design df         =         50
                                                F(   3,     48)   =      29.96
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0720

------------------------------------------------------------------------------
             |             Linearized
      weight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      height |   .4122218   .0577173     7.14   0.000     .2962932    .5281503
             |
        race |
      Black  |   13.97939   3.542194     3.95   0.000      6.86468    21.09409
```

```
      Other  |  -6.629714   2.798699    -2.37   0.022    -12.25107   -1.008361
             |
       _cons |  -26.49653   23.63431    -1.12   0.268    -73.96743    20.97438
------------------------------------------------------------------------------

.
.
.
. *************************************************************************************
. * Question 2: Do X1 and X2 have significant, but different relations with Y?
. * Question 3: Do the effects of X1 and X2 cancel each out?
. *************************************************************************************
.
.
.         ***********************************************************************
.         * 2.1 Modifying Mdoel 1.1.3
.         ************************************************************************
.
.         reg weight height i.race if sex ==2 [pw=sampwgt], cluster(school)
(sum of wgt is 4,151,978.6416772)

Linear regression                               Number of obs     =      2,133
                                                F(3, 9)           =      29.63
                                                Prob > F          =     0.0001
                                                R-squared         =     0.0720
                                                Root MSE          =      30.32

                                (Std. Err. adjusted for 10 clusters in school)
------------------------------------------------------------------------------
             |               Robust
      weight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      height |   .4122218   .0613759     6.72   0.000     .2733797    .5510638
             |
        race |
      Black  |   13.97939   3.113373     4.49   0.002     6.936449    21.02233
      Other  |  -6.629714   2.649844    -2.50   0.034    -12.62408   -.6353504
             |
       _cons |  -26.49653   25.33738    -1.05   0.323    -83.81367    30.82062
------------------------------------------------------------------------------

.         reg,  coeflegend

Linear regression                               Number of obs     =      2,133
                                                F(3, 9)           =      29.63
                                                Prob > F          =     0.0001
                                                R-squared         =     0.0720
                                                Root MSE          =      30.32

                                (Std. Err. adjusted for 10 clusters in school)
------------------------------------------------------------------------------
      weight |      Coef.  Legend
-------------+----------------------------------------------------------------
      height |   .4122218  _b[height]
             |
        race |
      Black  |   13.97939  _b[2.race]
      Other  |  -6.629714  _b[3.race]
             |
       _cons |  -26.49653  _b[_cons]
------------------------------------------------------------------------------

.
```

```
.         *testing the equlaity of two regression coefficients
.         test _b[2.race] = _b[3.race]

 ( 1)  2.race - 3.race = 0

       F(  1,     9) =   16.27
            Prob > F =    0.0030

.
.         *testing the hypothesis that variables cancel out each other
.         test _b[2.race] =-_b[3.race]

 ( 1)  2.race + 3.race = 0

       F(  1,     9) =    7.37
            Prob > F =    0.0238

.
.
.         *********************************************************************
.         * 2.2 Regression Analysis Using Results from Model 1.2.3.
.         *********************************************************************
.
.         svy, subpop(if sex==2): reg weight height i.race
(running regress on estimation sample)

Survey: Linear regression

Number of strata   =        50                  Number of obs     =      4,071
Number of PSUs     =       100                  Population size   =  8,000,000
                                                Subpop. no. obs   =      2,133
                                                Subpop. size      =  4,151,979
                                                Design df         =         50
                                                F(   3,     48)   =      30.05
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0720

------------------------------------------------------------------------------
             |             Linearized
      weight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      height |   .4122218   .0576472     7.15   0.000     .2964339    .5280096
             |
        race |
      Black  |   13.97939   3.538247     3.95   0.000      6.87261    21.08616
      Other  |  -6.629714   2.787416    -2.38   0.021     -12.2284   -1.031025
             |
       _cons |  -26.49653    23.6053    -1.12   0.267    -73.90917    20.91612
------------------------------------------------------------------------------

.         reg,  coeflegend

Survey: Linear regression

Number of strata   =        50                  Number of obs     =      4,071
Number of PSUs     =       100                  Population size   =  8,000,000
                                                Subpop. no. obs   =      2,133
                                                Subpop. size      =  4,151,979
                                                Design df         =         50
                                                F(   3,     48)   =      30.05
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0720
```

```
------------------------------------------------------------------------------
      weight |      Coef.  Legend
-------------+----------------------------------------------------------------
      height |   .4122218  _b[height]
             |
        race |
      Black  |   13.97939  _b[2.race]
      Other  |  -6.629714  _b[3.race]
             |
       _cons |  -26.49653  _b[_cons]
------------------------------------------------------------------------------

.
.         *testing the equlaity of two regression coefficients
.         test _b[2.race] = _b[3.race]

Adjusted Wald test

 ( 1)  2.race - 3.race = 0

       F(  1,    50) =   23.75
            Prob > F =    0.0000

.
.         *testing the hypothesis that variables cancel out each other
.         test _b[2.race] =-_b[3.race]

Adjusted Wald test

 ( 1)  2.race + 3.race = 0

       F(  1,    50) =    2.38
            Prob > F =    0.1292

.
.
.
. ********************************************************************************
. * Questions 4 : Does the relation between X1 and Y change with the levels of Z?
. *********************************************************************************
.
.
.         ************************************************************************
.         * 4.1 A three-way interactdion: Modify the model 2.1
.         ************************************************************************
.
.         reg weight c.height##i.race##i.sex [pw=sampwgt], cluster(school)
(sum of wgt is 8,000,000)

Linear regression                               Number of obs     =      4,071
                                                F(8, 9)           =          .
                                                Prob > F          =          .
                                                R-squared         =     0.3057
                                                Root MSE          =     28.633

                                    (Std. Err. adjusted for 10 clusters in school)
-----------------------------------------------------------------------------------
                  |               Robust
           weight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+----------------------------------------------------------------
           height |    .740487   .0528668    14.01   0.000     .6208939    .8600801
                  |
             race |
```

```
           Black  |  -14.95026   62.69483    -0.24   0.817    -156.7758    126.8753
           Other  |  -163.9238   47.83865    -3.43   0.008    -272.1424    -55.7053
                  |
    race#c.height |
           Black  |   .0288018   .1417587     0.20   0.844    -.2918788    .3494823
           Other  |   .3403916   .1064657     3.20   0.011     .0995494    .5812338
                  |
              sex |
          female  |   136.4661   42.54978     3.21   0.011     40.21179    232.7204
                  |
     sex#c.height |
          female  |  -.3405318   .1005027    -3.39   0.008    -.5678846   -.1131789
                  |
         race#sex |
    Black#female  |   8.525301   104.1438     0.08   0.937    -227.0643    244.1149
    Other#female  |   13.26205   178.4405     0.07   0.942    -390.3983    416.9224
                  |
race#sex#c.height |
    Black#female  |   .0202906   .2453719     0.08   0.936    -.5347792    .5753605
    Other#female  |   .0183652    .437561     0.04   0.967    -.9714665    1.008197
                  |
            _cons |  -157.8568   23.37035    -6.75   0.000    -210.7243   -104.9894
-----------------------------------------------------------------------------------

.         reg,  coeflegend

Linear regression                               Number of obs     =      4,071
                                                F(8, 9)           =          .
                                                Prob > F          =          .
                                                R-squared         =     0.3057
                                                Root MSE          =     28.633

                                     (Std. Err. adjusted for 10 clusters in school)
-----------------------------------------------------------------------------------
           weight |      Coef.  Legend
------------------+----------------------------------------------------------------
           height |    .740487  _b[height]
                  |
             race |
           Black  |  -14.95026  _b[2.race]
           Other  |  -163.9238  _b[3.race]
                  |
    race#c.height |
           Black  |   .0288018  _b[2.race#c.height]
           Other  |   .3403916  _b[3.race#c.height]
                  |
              sex |
          female  |   136.4661  _b[2.sex]
                  |
     sex#c.height |
          female  |  -.3405318  _b[2.sex#c.height]
                  |
         race#sex |
    Black#female  |   8.525301  _b[2.race#2.sex]
    Other#female  |   13.26205  _b[3.race#2.sex]
                  |
race#sex#c.height |
    Black#female  |   .0202906  _b[2.race#2.sex#c.height]
    Other#female  |   .0183652  _b[3.race#2.sex#c.height]
                  |
            _cons |  -157.8568  _b[_cons]
-----------------------------------------------------------------------------------
```

```
.
.         * test the coefficients of two-way interactions
.           test _b[2.race#c.height] = _b[3.race#c.height]

 ( 1)  2.race#c.height - 3.race#c.height = 0

       F(  1,     9) =    5.14
            Prob > F =    0.0496

.           test _b[2.race#c.height] = -_b[3.race#c.height]

 ( 1)  2.race#c.height + 3.race#c.height = 0

       F(  1,     9) =    3.10
            Prob > F =    0.1122

.
.         * test the coefficients of three-way interactions
.           test _b[2.race#2.sex#c.height] = _b[3.race#2.sex#c.height]

 ( 1)  2.race#2.sex#c.height - 3.race#2.sex#c.height = 0

       F(  1,     9) =    0.00
            Prob > F =    0.9973

.           test _b[2.race#2.sex#c.height] = -_b[3.race#2.sex#c.height]

 ( 1)  2.race#2.sex#c.height + 3.race#2.sex#c.height = 0

       F(  1,     9) =    0.01
            Prob > F =    0.9329

.
.
.
.         ********************************************************************
.         * 4.2 A two-way interaction Modify the model 2.2.
.         ********************************************************************
.
.         svy: reg weight c.height##i.race##i.sex
(running regress on estimation sample)

Survey: Linear regression

Number of strata   =        50                  Number of obs     =      4,071
Number of PSUs     =       100                  Population size   =  8,000,000
                                                Design df         =         50
                                                F(  11,     40)   =      65.77
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3057

------------------------------------------------------------------------------
                  |             Linearized
           weight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
           height |    .740487   .0549802    13.47   0.000      .630056    .8509181
                  |
             race |
           Black  |  -14.95026    77.7808    -0.19   0.848    -171.1776    141.2771
           Other  |  -163.9238   72.88719    -2.25   0.029    -310.3221   -17.52561
                  |
    race#c.height |
           Black  |   .0288018   .1734996     0.17   0.869    -.3196824    .3772859
```

```
           Other  |   .3403916   .1630919     2.09   0.042     .0128119    .6679713
                  |
              sex |
          female  |   136.4661   33.61345     4.06   0.000      68.9515    203.9807
                  |
     sex#c.height |
          female  |  -.3405318   .0793713    -4.29   0.000    -.4999537   -.1811098
                  |
         race#sex |
    Black#female  |   8.525301   102.9227     0.08   0.934     -198.201    215.2516
    Other#female  |   13.26205   164.6361     0.08   0.936    -317.4193    343.9434
                  |
race#sex#c.height |
    Black#female  |   .0202906   .2392789     0.08   0.933    -.4603152    .5008964
    Other#female  |   .0183652   .3981457     0.05   0.963    -.7813339    .8180642
                  |
            _cons |  -157.8568   24.38347    -6.47   0.000    -206.8325   -108.8812
-----------------------------------------------------------------------------------

.         reg,  coeflegend

Survey: Linear regression

Number of strata   =        50                  Number of obs     =      4,071
Number of PSUs     =       100                  Population size   =  8,000,000
                                                Design df         =         50
                                                F(  11,     40)   =      65.77
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3057

-----------------------------------------------------------------------------------
           weight |      Coef.  Legend
------------------+----------------------------------------------------------------
           height |    .740487  _b[height]
                  |
             race |
           Black  |  -14.95026  _b[2.race]
           Other  |  -163.9238  _b[3.race]
                  |
    race#c.height |
           Black  |   .0288018  _b[2.race#c.height]
           Other  |   .3403916  _b[3.race#c.height]
                  |
              sex |
          female  |   136.4661  _b[2.sex]
                  |
     sex#c.height |
          female  |  -.3405318  _b[2.sex#c.height]
                  |
         race#sex |
    Black#female  |   8.525301  _b[2.race#2.sex]
    Other#female  |   13.26205  _b[3.race#2.sex]
                  |
race#sex#c.height |
    Black#female  |   .0202906  _b[2.race#2.sex#c.height]
    Other#female  |   .0183652  _b[3.race#2.sex#c.height]
                  |
            _cons |  -157.8568  _b[_cons]
-----------------------------------------------------------------------------------

.
.         * test the coefficients of two-way interactions
.           test _b[2.race#c.height] = _b[3.race#c.height]
```

```
Adjusted Wald test

 ( 1)  2.race#c.height - 3.race#c.height = 0

       F(  1,    50) =    1.89
            Prob > F =    0.1749

.           test _b[2.race#c.height] = -_b[3.race#c.height]

Adjusted Wald test

 ( 1)  2.race#c.height + 3.race#c.height = 0

       F(  1,    50) =    2.19
            Prob > F =    0.1449

.
.         * test the coefficients of three-way interactions
.           test _b[2.race#2.sex#c.height] = _b[3.race#2.sex#c.height]

Adjusted Wald test

 ( 1)  2.race#2.sex#c.height - 3.race#2.sex#c.height = 0

       F(  1,    50) =    0.00
            Prob > F =    0.9966

.           test _b[2.race#2.sex#c.height] = -_b[3.race#2.sex#c.height]

Adjusted Wald test

 ( 1)  2.race#2.sex#c.height + 3.race#2.sex#c.height = 0

       F(  1,    50) =    0.01
            Prob > F =    0.9358

.
.
. ********************************************************************************
. * Questions 5: testing the total effect of a variable X with the margins command and complex survy data
. *********************************************************************************
.
.         * test the total effect of sex
.         svy: reg weight c.height##i.race##i.sex
(running regress on estimation sample)

Survey: Linear regression

Number of strata   =        50                  Number of obs     =      4,071
Number of PSUs     =       100                  Population size   =  8,000,000
                                                Design df         =         50
                                                F(  11,     40)   =      65.77
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3057

------------------------------------------------------------------------------------
                  |             Linearized
           weight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
           height |    .740487   .0549802    13.47   0.000      .630056    .8509181
                  |
             race |
```

```
           Black  |  -14.95026    77.7808    -0.19   0.848    -171.1776    141.2771
           Other  |  -163.9238   72.88719    -2.25   0.029    -310.3221   -17.52561
                  |
    race#c.height |
           Black  |   .0288018   .1734996     0.17   0.869    -.3196824    .3772859
           Other  |   .3403916   .1630919     2.09   0.042     .0128119    .6679713
                  |
              sex |
          female  |   136.4661   33.61345     4.06   0.000      68.9515    203.9807
                  |
     sex#c.height |
          female  |  -.3405318   .0793713    -4.29   0.000    -.4999537   -.1811098
                  |
         race#sex |
    Black#female  |   8.525301   102.9227     0.08   0.934     -198.201    215.2516
    Other#female  |   13.26205   164.6361     0.08   0.936    -317.4193    343.9434
                  |
race#sex#c.height |
    Black#female  |   .0202906   .2392789     0.08   0.933    -.4603152    .5008964
    Other#female  |   .0183652   .3981457     0.05   0.963    -.7813339    .8180642
                  |
            _cons |  -157.8568   24.38347    -6.47   0.000    -206.8325   -108.8812
------------------------------------------------------------------------------------

.
.         * The total effect of sex for the whole sample
.         margins, by(sex)

Predictive margins                              Number of obs     =      4,071
Model VCE    : Linearized

Expression   : Linear prediction, predict()
over         : sex

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |
       male  |   175.4809   .9271741   189.26   0.000     173.6186    177.3432
     female  |    146.204   .8358546   174.92   0.000     144.5251    147.8829
------------------------------------------------------------------------------

.         marginsplot, by(sex)

  Variables that uniquely identify margins: sex

.
.         margins i.sex,  at(height=(355(50)515)) atmeans

Adjusted predictions                            Number of obs     =      4,071
Model VCE    : Linearized

Expression   : Linear prediction, predict()

1._at        : height          =         355
               1.race          =    .8785304 (mean)
               2.race          =     .097057 (mean)
               3.race          =    .0244125 (mean)
               1.sex           =    .4810027 (mean)
               2.sex           =    .5189973 (mean)

2._at        : height          =         405
```

```
               1.race          =    .8785304 (mean)
               2.race          =     .097057 (mean)
               3.race          =    .0244125 (mean)
               1.sex           =    .4810027 (mean)
               2.sex           =    .5189973 (mean)

3._at        : height          =         455
               1.race          =    .8785304 (mean)
               2.race          =     .097057 (mean)
               3.race          =    .0244125 (mean)
               1.sex           =    .4810027 (mean)
               2.sex           =    .5189973 (mean)

4._at        : height          =         505
               1.race          =    .8785304 (mean)
               2.race          =     .097057 (mean)
               3.race          =    .0244125 (mean)
               1.sex           =    .4810027 (mean)
               2.sex           =    .5189973 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     _at#sex |
     1#male  |   103.5056   4.463539    23.19   0.000     94.54031    112.4709
   1#female  |   121.0924   3.247045    37.29   0.000     114.5705    127.6143
     2#male  |   141.0852   2.092828    67.41   0.000     136.8816    145.2888
   2#female  |   141.7663   .8083906   175.37   0.000     140.1426      143.39
     3#male  |   178.6648   1.030582   173.36   0.000     176.5948    180.7348
   3#female  |   162.4402   2.673706    60.75   0.000     157.0699    167.8105
     4#male  |   216.2444   3.157481    68.49   0.000     209.9024    222.5864
   4#female  |   183.1141   5.479865    33.42   0.000     172.1075    194.1208
------------------------------------------------------------------------------

.         marginsplot, yline(0)

  Variables that uniquely identify margins: height sex

.
.
.         margins i.sex,  at(height=(355(50)515)) at(race ==1)

Predictive margins                              Number of obs     =      4,071
Model VCE    : Linearized

Expression   : Linear prediction, predict()

1._at        : height          =         355

2._at        : height          =         405

3._at        : height          =         455

4._at        : height          =         505

5._at        : race            =           1

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     _at#sex |
```

```
     1#male  |   103.5056   4.463539    23.19   0.000     94.54031    112.4709
   1#female  |   121.0924   3.247045    37.29   0.000     114.5705    127.6143
     2#male  |   141.0852   2.092828    67.41   0.000     136.8816    145.2888
   2#female  |   141.7663   .8083906   175.37   0.000     140.1426      143.39
     3#male  |   178.6648   1.030582   173.36   0.000     176.5948    180.7348
   3#female  |   162.4402   2.673706    60.75   0.000     157.0699    167.8105
     4#male  |   216.2444   3.157481    68.49   0.000     209.9024    222.5864
   4#female  |   183.1141   5.479865    33.42   0.000     172.1075    194.1208
     5#male  |    162.507   1.048197   155.03   0.000     160.4016    164.6124
   5#female  |   151.6457   1.688154    89.83   0.000     148.2549    155.0364
------------------------------------------------------------------------------

.         marginsplot, yline(0) saving("D:\Jason\workshop\regression analysis 2022\graph_white.gph", replace)

  Variables that uniquely identify margins: height sex
  Multiple at() options specified:
      _atoption=1: height=(355(50)515)
      _atoption=2: race ==1
(file D:\Jason\workshop\regression analysis 2022\graph_white.gph saved)

.
.
.         margins i.sex,  at(height=(355(50)515)) at(race ==2)

Predictive margins                              Number of obs     =      4,071
Model VCE    : Linearized

Expression   : Linear prediction, predict()

1._at        : height          =         355

2._at        : height          =         405

3._at        : height          =         455

4._at        : height          =         505

5._at        : race            =           2

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     _at#sex |
     1#male  |   103.5056   4.463539    23.19   0.000     94.54031    112.4709
   1#female  |   121.0924   3.247045    37.29   0.000     114.5705    127.6143
     2#male  |   141.0852   2.092828    67.41   0.000     136.8816    145.2888
   2#female  |   141.7663   .8083906   175.37   0.000     140.1426      143.39
     3#male  |   178.6648   1.030582   173.36   0.000     176.5948    180.7348
   3#female  |   162.4402   2.673706    60.75   0.000     157.0699    167.8105
     4#male  |   216.2444   3.157481    68.49   0.000     209.9024    222.5864
   4#female  |   183.1141   5.479865    33.42   0.000     172.1075    194.1208
     5#male  |   160.0175   3.840411    41.67   0.000     152.3038    167.7312
   5#female  |     166.46   4.283372    38.86   0.000     157.8566    175.0634
------------------------------------------------------------------------------

.        marginsplot, yline(0) saving("D:\Jason\workshop\regression analysis 2022\graph_black.gph", replace)

  Variables that uniquely identify margins: height sex
  Multiple at() options specified:
      _atoption=1: height=(355(50)515)
      _atoption=2: race ==2
(file D:\Jason\workshop\regression analysis 2022\graph_black.gph saved)
```

```
.
.         margins i.sex,  at(height=(355(50)515)) at(race ==3)

Predictive margins                              Number of obs     =      4,071
Model VCE    : Linearized

Expression   : Linear prediction, predict()

1._at        : height          =         355

2._at        : height          =         405

3._at        : height          =         455

4._at        : height          =         505

5._at        : race            =           3

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     _at#sex |
     1#male  |   103.5056   4.463539    23.19   0.000     94.54031    112.4709
   1#female  |   121.0924   3.247045    37.29   0.000     114.5705    127.6143
     2#male  |   141.0852   2.092828    67.41   0.000     136.8816    145.2888
   2#female  |   141.7663   .8083906   175.37   0.000     140.1426      143.39
     3#male  |   178.6648   1.030582   173.36   0.000     176.5948    180.7348
   3#female  |   162.4402   2.673706    60.75   0.000     157.0699    167.8105
     4#male  |   216.2444   3.157481    68.49   0.000     209.9024    222.5864
   4#female  |   183.1141   5.479865    33.42   0.000     172.1075    194.1208
     5#male  |     145.85   4.438492    32.86   0.000      136.935    154.7649
   5#female  |   156.1962   10.10268    15.46   0.000     135.9043     176.488
------------------------------------------------------------------------------

.        marginsplot, yline(0) saving("D:\Jason\workshop\regression analysis 2022\graph_other.gph", replace)

  Variables that uniquely identify margins: height sex
  Multiple at() options specified:
      _atoption=1: height=(355(50)515)
      _atoption=2: race ==3
(file D:\Jason\workshop\regression analysis 2022\graph_other.gph saved)

.
.
.         margins i.sex, atmeans

Adjusted predictions                            Number of obs     =      4,071
Model VCE    : Linearized

Expression   : Linear prediction, predict()
at           : height          =    432.6394 (mean)
               1.race          =    .8785304 (mean)
               2.race          =     .097057 (mean)
               3.race          =    .0244125 (mean)
               1.sex           =    .4810027 (mean)
               2.sex           =    .5189973 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
```

```
        sex |
       male |   161.8587   1.017217   159.12   0.000     159.8156    163.9019
     female |   153.1946   1.489519   102.85   0.000     150.2028    156.1864
------------------------------------------------------------------------------

.         margins, dydx(i.sex) atmeans

Conditional marginal effects                    Number of obs     =      4,071
Model VCE    : Linearized

Expression   : Linear prediction, predict()
dy/dx w.r.t. : 2.sex
at           : height          =    432.6394 (mean)
               1.race          =    .8785304 (mean)
               2.race          =     .097057 (mean)
               3.race          =    .0244125 (mean)
               1.sex           =    .4810027 (mean)
               2.sex           =    .5189973 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |
     female  |  -8.664163   1.972424    -4.39   0.000    -12.62589   -4.702433
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.

.         margins, dydx(i.sex i.race) atmeans

Conditional marginal effects                    Number of obs     =      4,071
Model VCE    : Linearized

Expression   : Linear prediction, predict()
dy/dx w.r.t. : 2.race 3.race 2.sex
at           : height          =    432.6394 (mean)
               1.race          =    .8785304 (mean)
               2.race          =     .097057 (mean)
               3.race          =    .0244125 (mean)
               1.sex           =    .4810027 (mean)
               2.sex           =    .5189973 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        race |
      Black  |   6.491153   3.199031     2.03   0.048     .0657108     12.9166
      Other  |  -5.650381   6.262037    -0.90   0.371    -18.22805     6.92729
             |
         sex |
     female  |  -8.664163   1.972424    -4.39   0.000    -12.62589   -4.702433
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.

.
.
. log close
      name:  <unnamed>
       log:  D:\Jason\workshop\regression analysis 2022\regression2.log
  log type:  text
 closed on:   3 Oct 2022, 10:52:43
-------------------------------------------------------------------------------------------------------------------------------------------------------------------
```