SEs and 95% CI for linear combination of the estimates in STATA

In Stata we can estimate the mean, the SE and the 95% CI of a linear combination of the parameters by using the command `lincom` after the model fit.

For example, to estimate mean plasma of retinol in men with infrequent vitamin use we need to estimate the combination: Comb1=b_{men}+b_{infreq}+b_{men_infreq}+_cons.

If the model has been defined as: reg retplasm men freq infeq menfreq meninfr

We can get the estimate, the SE and the 95% CI of the parameter Comb1 in Stata by the command:

```
Lincom men+infreq+men_infreq+_cons
```

after fitting the model.

If the model has been defined using the `xi` command , we must trace and use the appropriate dummy variables (`I*`) in Stata's "variables" list.

For example:

```
. char sex[omit] 2
. char vituse[omit] 3
. xi:reg  retplasm i.sex*i.vituse
i.sex             _Isex_1-2           (naturally coded; _Isex_2 omitted)
i.vituse          _Ivituse_1-3        (naturally coded; _Ivituse_3 omitted)
i.sex*i.vituse    _IsexXvit_#_#       (coded as above)

      Source |       SS           df       MS      Number of obs   =       314
-------------+----------------------------------   F(5, 308)       =      4.03
       Model |  836388.279         5  167277.656   Prob > F        =    0.0015
    Residual |  12783793.6       308  41505.8233   R-squared       =    0.0614
-------------+----------------------------------   Adj R-squared   =    0.0462
       Total |  13620181.9       313   43514.958   Root MSE        =    203.73


------------------------------------------------------------------------------
     retplasm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
       _Isex_1 |   166.3468   47.76693     3.48   0.001     72.35603    260.3376
    _Ivituse_1 |   33.46968   29.28935     1.14   0.254    -24.16285    91.10221
    _Ivituse_2 |   39.49589   31.87656     1.24   0.216    -23.22749    102.2193
_IsexXvit_1_1 |  -11.72721   76.51943    -0.15   0.878    -162.2942    138.8398
_IsexXvit_1_2 |  -255.6611   105.4603    -2.42   0.016     -463.175   -48.14725
         _cons |   563.2184   21.84213    25.79   0.000     520.2397    606.1971
------------------------------------------------------------------------------
```

Stata has now created a dummy variable for sex (`Isex_1`), two dummy variables for vitamine use (`Ivituse_1 Ivituse_2`) and two dummy variables for their respective interactions (`IsexXvit_1_1 IsexXvit_1_2`). Notice the reference levels for these two categorical variables as defined by the "`char varname[omit] #`" commands. Thus if we want to estimate mean retinol plasma levels for men who use vitamins infrequently (sex=2 & vituse=2) we can use the following command:

```
. lincom  Isex_1+ Ivituse_2+ IsexXvit_1_2+_cons

 ( 1)  Isex_1 + Ivituse_2 + IsexXvit_1_2 + _cons = 0.0


------------------------------------------------------------------------------
retplasm |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     (1) |      513.4   91.11073      5.635   0.000       334.1218    692.6782
```

These results are based on standard probability theory [Var(A+B)=Var(A)+Var(B)+2*Cov(A, B)], and can be obtained "by hand" using some simple matrix algebra and taking advantage of the fact that Stata saves the vector of coefficients as e(b) and their respective variance-covariance matrix as e(V) after each model estimation command:

We can print these two matrices using the "`matrix list`" command.

```
. mat li e(b)
e(b)[1,6]
         _Isex_1    _Ivituse_1    _Ivituse_2  _IsexXvit_~1  _IsexXvit_~2         _cons
y1     166.34683     33.469683     39.495895    -11.727208    -255.66111     563.21839


. mat li e(V)
symmetric e(V)[6,6]
                  _Isex_1    _Ivituse_1    _Ivituse_2  _IsexXvit_~1  _IsexXvit_~2         _cons
     _Isex_1    2281.6794
  _Ivituse_1    477.07843     857.8658
  _Ivituse_2    477.07843    477.07843     1016.1151
_IsexXvit_~1   -2281.6794    -857.8658    -477.07843     5855.2225
_IsexXvit_~2   -2281.6794   -477.07843    -1016.1151     2281.6794     11121.881
       _cons   -477.07843   -477.07843    -477.07843     477.07843     477.07843     477.07843
```

Check the relation between the output of the `regress` command and the elements of those two matrices.

We can now estimate any linear combination of the coefficients of the model and its standard error (or variance). The calculations can be simplified by defining a vector C of length 6 (our model has 6 coefficients) with all its elements equal to 0 except those who correspond to the coefficients that we want to include in the linear combinations (constraint). These elements must be set equal to 1. For example if we want to replicate the `lincom` command as presented above our C vector must be C=(1, 0, 1, 0, 1, 1).

```
. matrix C=(1,0,1,0,1,1)
```

By multiplying the 1x6 e(b) matrix with the transpose of the 1x6 C vector we have a scalar (1x1) which is equal with the sum of the appropriate model coefficients

(`Isex_1 + Ivituse_1 + Ivituse_2 + IsexXvit_1_1 + IsexXvit_1_2 + _cons`)

```
. matrix comb1=e(b)*C'
. mat li comb1
symmetric comb1[1,1]
       r1
y1  513.4
```

The variance (and the standard error of this estimate) can be easily obtained by the following matrix product Var(comb1)=C*e(V)*C^T

```
. matrix Vcomb1=C*e(V)*C'
. matrix li Vcomb1

symmetric Vcomb1[1,1]
           r1
r1  8301.1647
```

The standard error of comb1 is just the square root of its variance and can be obtained by the following two ways with the second being more general:

```
. local vcomb1=Vcomb1[1,1]
. di sqrt(`vcomb1')
91.110727
```

or

```
. mat SEcomb1=cholesky(Vcomb1)
. mat li SEcomb1
symmetric SEcomb1[1,1]
           r1
r1  91.110727
```

Notice that in the first approach we used Stata's local macros to extract the element of Vcomb1 as a scalar in order to calculate its square root, while in the second we used the Cholesky matrix decomposition which in the 1-dimensional case is equivalent to the square root. The second approach is more general while it allows for multiple linear combinations' estimation along with their standard errors by just replacing the 1xp C vector (p= # of coefficients in the model) with the appropriate mxp matrix where m is the number of linear combinations to be estimated. The C matrix must be formed in the same manner as the C vector in our example*.

The 95% confidence interval can be obtained by adding or substracting the product of the standard error with the $t_{0.975,df}$ value (remember that we have an OLS estimation case):

```
. *Lower 95% Limit
. di 513.4 -invttail(308,0.025)*(91.11073)
334.12179
. *Upper 95% Limit
. di 513.4 +invttail(308,0.025)*(91.11073)
692.67821
```

Notice the number of degrees of freedom (df=308) and the p argument (0.025) in the `invttail(df,p)` Stata's function. In case of ML estimation we must use the $z_{0.975}$ value instead of the $t_{0.975,df}$ one.

The procedure described above is a general method for estimating linear combinations of a model's coefficients and can be very useful especially in statistical packages lacking the "`lincom`" Stata's utility.

*We can estimate two linear combinations (men infrequent and men frequent users) simultaneously as follows:

```
. matrix C=(1,0,1,0,1,1\1,1,0,1,0,1)
. matrix comb1=e(b)*C'
. mat li comb1
comb1[1,2]
           r1         r2
y1      513.4  751.30769
. matrix Vcomb1=C*e(V)*C'
. di sqrt(Vcomb1[1,1])
91.110727
. di sqrt(Vcomb1[2,2])
56.504475
```

Compare with the following `lincom` commands:

```
. lincom  _Isex_1+ _Ivituse_2+ _IsexXvit_1_2+ _cons
 ( 1)  _Isex_1 + _Ivituse_2 + _IsexXvit_1_2 + _cons = 0
------------------------------------------------------------------------------
    retplasm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |      513.4   91.11073     5.63   0.000     334.1218    692.6782
------------------------------------------------------------------------------

. lincom  _Isex_1+ _Ivituse_1+ _IsexXvit_1_1+ _cons
 ( 1)  _Isex_1 + _Ivituse_1 + _IsexXvit_1_1 + _cons = 0
------------------------------------------------------------------------------
    retplasm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   751.3077   56.50447    13.30   0.000     640.1241    862.4913
------------------------------------------------------------------------------
```