Interactions made easy

André Charlett

Neville Q Verlander

Health Protection Agency Centre for Infections

Motivation

Scientific staff within institute using Stata to fit many types of regression models using a variety of approaches

GLIM macros

lrtest rather tedious

Some issues

Somewhat “tiresome” to always remember to use the i. prefix for factors

Some mindless modelling via the *sw* command, especially with dummy variables

Wald test used rather the *lrtest*

What was required?

An automated program to allow users to specify a regression model which would return an appropriate hypothesis test for each term in the model.

$$y = \phi\left(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \ldots + \beta_i x_i\right) + \varepsilon$$

What is available?

test/testparm

```
Logistic regression                               Number of obs   =       6105
                                                  LR chi2(6)      =     119.58
                                                  Prob > chi2     =     0.0000
Log likelihood = -1416.9675                       Pseudo R2       =     0.0405

------------------------------------------------------------------------------
         ssi | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      _Iag_2 |   .9469958   .1465584    -0.35   0.725     .6992216     1.28257
      _Iag_3 |   1.192908   .1820896     1.16   0.248     .8844565     1.60893
      _Iag_4 |   1.481345   .2224766     2.62   0.009     1.103616    1.988357
_Iasascore_2 |   1.583486   .1946745     3.74   0.000      1.24442    2.014936
      _Iwc_1 |   6.081108   1.624676     6.76   0.000      3.60221    10.2658
-------------------------------------------------------------------------------

. testparm  _Iag_2 _Iag_3 _Iag_4

 ( 1)  _Iag_2 = 0
 ( 2)  _Iag_3 = 0
 ( 3)  _Iag_4 = 0

           chi2(  3) =   11.44
         Prob > chi2 =    0.0096
```

What is available?

lrtest

```
. quietly xi: logistic ssi i.ag i.asa i.wc duration

. est store full

. quietly xi: logistic ssi i.asa i.wc duration

. lrtest full

likelihood-ratio test                                  LR chi2(3)  =     11.24
(Assumption: . nested in full)                         Prob > chi2 =    0.0105
```

What is available?

lrdrop1 (STB-54: sg133)

```
. lrdrop1
Likelihood Ratio Tests: drop 1 term
logistic regression
number of obs = 6105
------------------------------------------------------------------------
     ssi     Df      Chi2      P>Chi2    -2*log ll   Res. Df   AIC
------------------------------------------------------------------------
Original Model                           2833.94     6098    2847.94
   -Iag*      3     11.24      0.0105    2845.18     6095    2853.18
-asascore     1     14.89      0.0001    2848.82     6097    2860.82
     -wc      1     35.29      0.0000    2869.23     6097    2881.23
-duration     1     55.10      0.0000    2889.03     6097    2901.03
------------------------------------------------------------------------
Terms dropped one at a time in turn.
```

A bit fiddly to get to work properly (xi_6)

anova

anova can do what is required

```
. anova  mpg  rep78 weight*length  weight  length foreign, cont( weight length)

                           Number of obs =      69     R-squared     =  0.7300
                           Root MSE      = 3.24516     Adj R-squared =  0.6940

                  Source |  Partial SS    df       MS           F     Prob > F
           --------------+----------------------------------------------------
                   Model |  1708.33912     8  213.542391      20.28     0.0000
                         |
                   rep78 |  101.826828     4  25.4567069       2.42     0.0584
           weight*length |  30.1591926     1  30.1591926       2.86     0.0958
                  weight |  49.2790347     1  49.2790347       4.68     0.0345
                  length |  78.8054893     1  78.8054893       7.48     0.0082
                 foreign |  65.8891197     1  65.8891197       6.26     0.0151
                         |
                Residual |  631.863774    60  10.5310629
           --------------+----------------------------------------------------
                   Total |   2340.2029    68  34.4147485
```

fitint command syntax

```
fitint regression_cmd yvar xvarlist [weight] [if exp]
[in range] [, factor(varlist) twoway(varlist [,varlist] )
noshow  regression_cmd_options ]
```

The *fitint* command

regression_cmd one of the following:

```
clogit
cloglog
cnreg
glm
logistic
logit
nbreg
ologit
oprobit
poisson
probit
regress
scobit
stcox
streg
tobit
```

N.B. *yvar* not required for stcox and streg commands but the data must be stset

The *fitint* command

- Generates variables using naming convention __X_Y for interactions between continuous variables.

- Looks for cluster(), robust, and noconstant options and will exit if detected.

- Some standard checks also done, e.g. factor list a subset of *xvarlist*

The *fitint* options

- noshow suppress the regression table output

- factor(varlist) define those xvarlist terms that are factors, analogous to category(varlist) option with *anova*

- twoway(varlist[,varlist]) defines those xvarlist terms for which two-way interactions are required. If more than two x-variables are listed then all possible two-way interactions are generated. When a comma is used to separate x-variable lists then all possible two-way interactions within each list are generated. e.g.

  twoway(A B C, D E) will produce the four interactions A*B, A*C, B*C, and D*E

Logistic regression example

```
fitint logistic ssi sg gender preopstay
typesurgery asascore wc durationoperation,
factor (sg gender typesurgery asascore wc)
twoway (gender wc, gender typesurgery)
```

```
i.sg              _Isg_1-5          (naturally coded; _Isg_1 omitted)
i.gender          _Igender_1-2      (naturally coded; _Igender_1 omitted)
i.typesurgery     _Itypesurge_1-2   (naturally coded; _Itypesurge_1 omitted)
i.asascore        _Iasascore_1-2    (naturally coded; _Iasascore_1 omitted)
i.wc              _Iwc_0-1          (naturally coded; _Iwc_0 omitted)
i.gender*i.wc     _IgenXwc_#_#      (coded as above)
i.gen~r*i.typ~y   _IgenXtyp_#_#     (coded as above)

note: _Igender_2 dropped due to collinearity
note: _Iwc_1 dropped due to collinearity
note: _Igender_2 dropped due to collinearity
note: _Itypesurge_2 dropped due to collinearity

Logistic regression                               Number of obs   =       5916
                                                  LR chi2(12)     =     286.28
                                                  Prob > chi2     =     0.0000
Log likelihood = -1243.3562                       Pseudo R2       =     0.1032

------------------------------------------------------------------------------
         ssi | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _Isg_2 |   .0858664   .0509368    -4.14   0.000     .0268461    .2746407
       _Isg_3 |    1.48855   .3977673     1.49   0.137     .8816718    2.513158
       _Isg_4 |   1.450785   .3622168     1.49   0.136     .8893734    2.366585
       _Isg_5 |   2.174301   .3075851     5.49   0.000     1.647804    2.869023
   _Igender_2 |   1.403811   .1861615     2.56   0.011     1.082504    1.820488
    preopstay |   1.369205   .1205059     3.57   0.000     1.152266    1.626987
_Itypesurg~2 |   1.954332   .3335914     3.93   0.000     1.398633    2.730818
_Iasascore_2 |   1.371105   .1761401     2.46   0.014      1.06591    1.763685
        _Iwc_1 |   6.725279   2.183002     5.87   0.000       3.5597    12.70596
durationop~n |   1.003782    .000524     7.23   0.000     1.002755    1.004809
_IgenXwc_2_1 |   .1076087   .0886077    -2.71   0.007     .0214263    .5404402
_IgenXtyp_~2 |   .4568698   .1517149    -2.36   0.018     .2383033    .8759007
------------------------------------------------------------------------------
```

```
-------------------------------------------------------------------------
Fitting and testing any interactions and any main effects not included
in interaction terms using the change in deviance from the full model
when each term is removed in turn to obtain the likelihood ratio chi
square statistic
-------------------------------------------------------------------------

Model summary
Number of observations used in estimation:    5916
Regression command:  logistic
Dependent variable:       ssi
Full model deviance:  2486.71
degrees of freedom:        13

-------------------------------------------------------------------------
      Term           |   Model deviance      Chi-square    df       P>Chi
-------------------+-----------------------------------------------------
       i.sg          |          2597.55          110.84     4      0.0000
     preopstay       |          2499.58           12.87     1      0.0003
    i.asascore       |          2492.99            6.28     1      0.0122
 durationoperation |          2535.08           48.37     1      0.0000
   i.gender*i.wc     |          2496.85           10.14     1      0.0015
 i.gender*i.typesurgery|        2492.73            6.02     1      0.0142
-------------------------------------------------------------------------
```

```
. fitint stcox  drug age, factor(drug) twoway ( drug age)

   failure _d:  died
   analysis time _t:  studytime

Cox regression -- Breslow method for ties

No. of subjects =           48                     Number of obs   =        48
No. of failures =           31
Time at risk    =          744
                                                   LR chi2(5)      =     36.91
Log likelihood  =   -81.456452                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    _Idrug_2 |   .0211826   .1002649    -0.81   0.415     1.98e-06    226.4762
         age |    1.11156   .0547048     2.15   0.032     1.009349    1.224121
    _Idrug_3 |   .2595464   1.469272    -0.24   0.812     3.94e-06    17092.31
 _IdruXage_2 |   1.037087   .0834552     0.45   0.651     .8857645    1.214261
 _IdruXage_3 |   .9712592   .0977057    -0.29   0.772     .7974563    1.182942
------------------------------------------------------------------------------


-------------------------------------------------------------------------
Fitting and testing any interactions and any main effects not included
in interaction terms using the change in deviance from the full model
when each term is removed in turn to obtain the likelihood ratio chi
square statistic
-------------------------------------------------------------------------

Model summary
Number of observations used in estimation:      48
Regression command:       cox
Dependent variable:        _t
Full model deviance:   162.91
degrees of freedom:         5

-------------------------------------------------------------------------
      Term         |   Model deviance      Chi-square    df      P>Chi
-------------------+-----------------------------------------------------
    i.drug*age     |           163.31            0.39     2     0.8219
-------------------------------------------------------------------------
```

```
. fitint regress mpg  weight length  foreign , factor( foreign) twoway ( weight length)
i.foreign         _Iforeign_0-1       (naturally coded; _Iforeign_0 omitted)

      Source |       SS       df       MS              Number of obs =      74
-------------+------------------------------           F(  4,    69) =   38.48
       Model |  1687.14169     4  421.785422           Prob > F      =  0.0000
    Residual |  756.317773    69  10.9611271           R-squared     =  0.6905
-------------+------------------------------           Adj R-squared =  0.6725
       Total |  2443.45946    73  33.4720474           Root MSE      =  3.3108

------------------------------------------------------------------------------
         mpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      weight |  -.0134239   .0048943    -2.74   0.008    -.0231877     -.00366
      length |  -.2044416   .0822498    -2.49   0.015    -.3685254   -.0403578
 _Iforeign_1 |  -2.054767   1.061208    -1.94   0.057    -4.171819    .0622845
       __4_3 |   .0000451   .0000231     1.95   0.055    -9.44e-07    .0000912
       _cons |     74.528   13.72006     5.43   0.000     47.15722    101.8988
------------------------------------------------------------------------------

------------------------------------------------------------------------
Fitting and testing any interactions and any main effects not included
in interaction terms using the ratio of the mean square error of each
term and the residual mean square error to obtain an F ratio statistic
------------------------------------------------------------------------

Model summary
Number of observations used in estimation:      74
Regression command:           regress
Dependent variable:               mpg
Residual MSE:                   10.96
degrees of freedom:                69

------------------------------------------------------------------------
      Term          |   Mean square    F ratio  df1    df2          P>F
--------------------+---------------------------------------------------
     i.foreign      |         41.09       3.75     1      69       0.0569
   weight*length    |         41.85       3.82     1      69       0.0548
------------------------------------------------------------------------
```

Further developments

- use *xi3* rather than *xi* to enable three-way interactions
- Care is required to ensure that tests of the coefficients of interaction terms are not used solely in non-linear models. Explore the use of *predictnl*, *inteff* (Norton *et al* SJ 4_2 pp 154-167), *postgr3* and *vibl* suite (Mitchell *et al* SJ 5_1 pp 64 – 82)