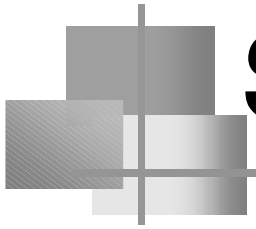
# Post-estimation commands for regression models for categorical & count outcomes

**Jeremy Freese**
University of Wisconsin-Madison
**J. Scott Long**
Indiana University

# Models for categorical and count outcomes

- Stata makes estimating these models easy
- Interpretation is more complicated
- Our **SPost** suite of commands is designed to facilitate interpretation and other tasks with these models
- Type `net search spost` to download



# SPost command: `fitstat`

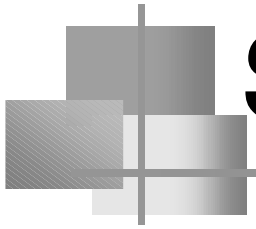
- Computes goodness-of-fit statistics
  - Both Pseudo-R2s and information measures (i.e., AIC and BIC)
- Can be used with saving() and using() options

Stata Results

```
. logit lfp k5 k618 age wc hc lwg inc, nolog

Logit estimates                                   Number of obs   =        753
                                                  LR chi2(7)      =     124.48
                                                  Prob > chi2     =     0.0000
Log likelihood = -452.63296                       Pseudo R2       =     0.1209

------------------------------------------------------------------------------
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          k5 |  -1.462913   .1970006    -7.43   0.000    -1.849027   -1.076799
        k618 |  -.0645707   .0680008    -0.95   0.342    -.1978499    .0687085
         age |  -.0628706   .0127831    -4.92   0.000    -.0879249   -.0378162
          wc |   .8072738   .2299799     3.51   0.000     .3565215    1.258026
          hc |   .1117336   .2060397     0.54   0.588    -.2920969     .515564
         lwg |   .6046931   .1508176     4.01   0.000     .3090961    .9002901
         inc |  -.0344464   .0082084    -4.20   0.000    -.0505346   -.0183583
       _cons |    3.18214   .6443751     4.94   0.000     1.919188    4.445092
------------------------------------------------------------------------------

. fitstat

Measures of Fit for logit of lfp

Log-Lik Intercept Only:     -514.873     Log-Lik Full Model:          -452.633
D(745):                      905.266     LR(7):                        124.480
                                         Prob > LR:                      0.000
McFadden's R2:                 0.121     McFadden's Adj R2:              0.105
Maximum Likelihood R2:         0.152     Cragg & Uhler's R2:             0.204
McKelvey and Zavoina's R2:     0.217     Efron's R2:                     0.155
Variance of y*:                4.203     Variance of error:              3.290
Count R2:                      0.693     Adj Count R2:                   0.289
AIC:                           1.223     AIC*n:                        921.266
BIC:                       -4029.663     BIC':                         -78.112

.
```

# Why are results from categorical and count models often difficult to interpret?

Nonlinearities mean that interpretation depends on the values of all independent variables.

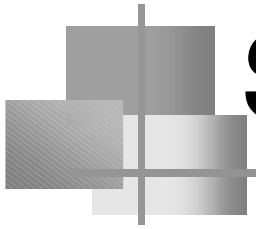


# SPost command: `prvalue`

- Produces predicted values for specified set of values of the independent variables
- Specific values are set with `x()` option
- All other values set with `rest()` option
- `save` and `dif` options to calculate differences between two sets of values

Stata Results

```
. logit lfp k5 k618 age wc hc lwg inc, nolog

Logit estimates                                   Number of obs   =        753
                                                  LR chi2(7)      =     124.48
                                                  Prob > chi2     =     0.0000
Log likelihood = -452.63296                       Pseudo R2       =     0.1209

------------------------------------------------------------------------------
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          k5 |  -1.462913   .1970006    -7.43   0.000    -1.849027   -1.076799
        k618 |  -.0645707   .0680008    -0.95   0.342    -.1978499    .0687085
         age |  -.0628706   .0127831    -4.92   0.000    -.0879249   -.0378162
          wc |   .8072738   .2299799     3.51   0.000     .3565215    1.258026
          hc |   .1117336   .2060397     0.54   0.588    -.2920969     .515564
         lwg |   .6046931   .1508176     4.01   0.000     .3090961    .9002901
         inc |  -.0344464   .0082084    -4.20   0.000    -.0505346   -.0183583
       _cons |    3.18214   .6443751     4.94   0.000     1.919188    4.445092
------------------------------------------------------------------------------

. prvalue, x(age=30 wc=1 inc=median) rest(mean)

logit: Predictions for lfp

  Pr(y=inLF|x):       0.8539   95% ci: (0.7789,0.9065)
  Pr(y=NotInLF|x):    0.1461   95% ci: (0.0935,0.2211)

           k5       k618        age         wc         hc        lwg
x=   .2377158  1.3532537         30          1  .39176627  1.0971148

          inc
x=  17.700001

.
```

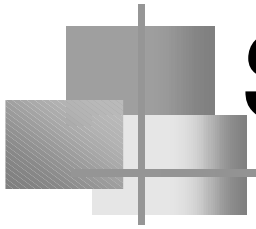


# SPost command: `prtab`

- Predicted probabilities for a cross-classification of 2-4 categorical independent variables
- Values of other variables specified by `x()` and `rest()`

Stata Results

```
. logit lfp k5 k618 age wc hc lwg inc, nolog

Logit estimates                                   Number of obs   =        753
                                                  LR chi2(7)      =     124.48
                                                  Prob > chi2     =     0.0000
Log likelihood = -452.63296                       Pseudo R2       =     0.1209

------------------------------------------------------------------------------
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          k5 |  -1.462913   .1970006    -7.43   0.000    -1.849027   -1.076799
        k618 |  -.0645707   .0680008    -0.95   0.342    -.1978499    .0687085
         age |  -.0628706   .0127831    -4.92   0.000    -.0879249   -.0378162
          wc |   .8072738   .2299799     3.51   0.000     .3565215    1.258026
          hc |   .1117836   .2060897     0.54   0.588    -.2920969     .515564
         lwg |   .6046931   .1508176     4.01   0.000     .3090961    .9002901
         inc |  -.0344464   .0082084    -4.20   0.000    -.0505346   -.0183583
       _cons |    3.18214   .6443751     4.94   0.000     1.919188    4.445092
------------------------------------------------------------------------------

. prtab k5 wc, x(k618=0) rest(mean)

logit: Predicted probabilities of positive outcome for lfp

-----------------------------------------
# kids <  |  Wife College:
6         |    1=yes 0=no
          |   NoCol   College
----------+------------------------------
        0 |  0.6275    0.7907
        1 |  0.2806    0.4665
        2 |  0.0828    0.1684
        3 |  0.0205    0.0448
-----------------------------------------

          k5       k618        age         wc         hc        lwg
x=  .2377158          0  42.537849  .2815405  .39176627  1.0971148

         inc
x=  20.128965
```

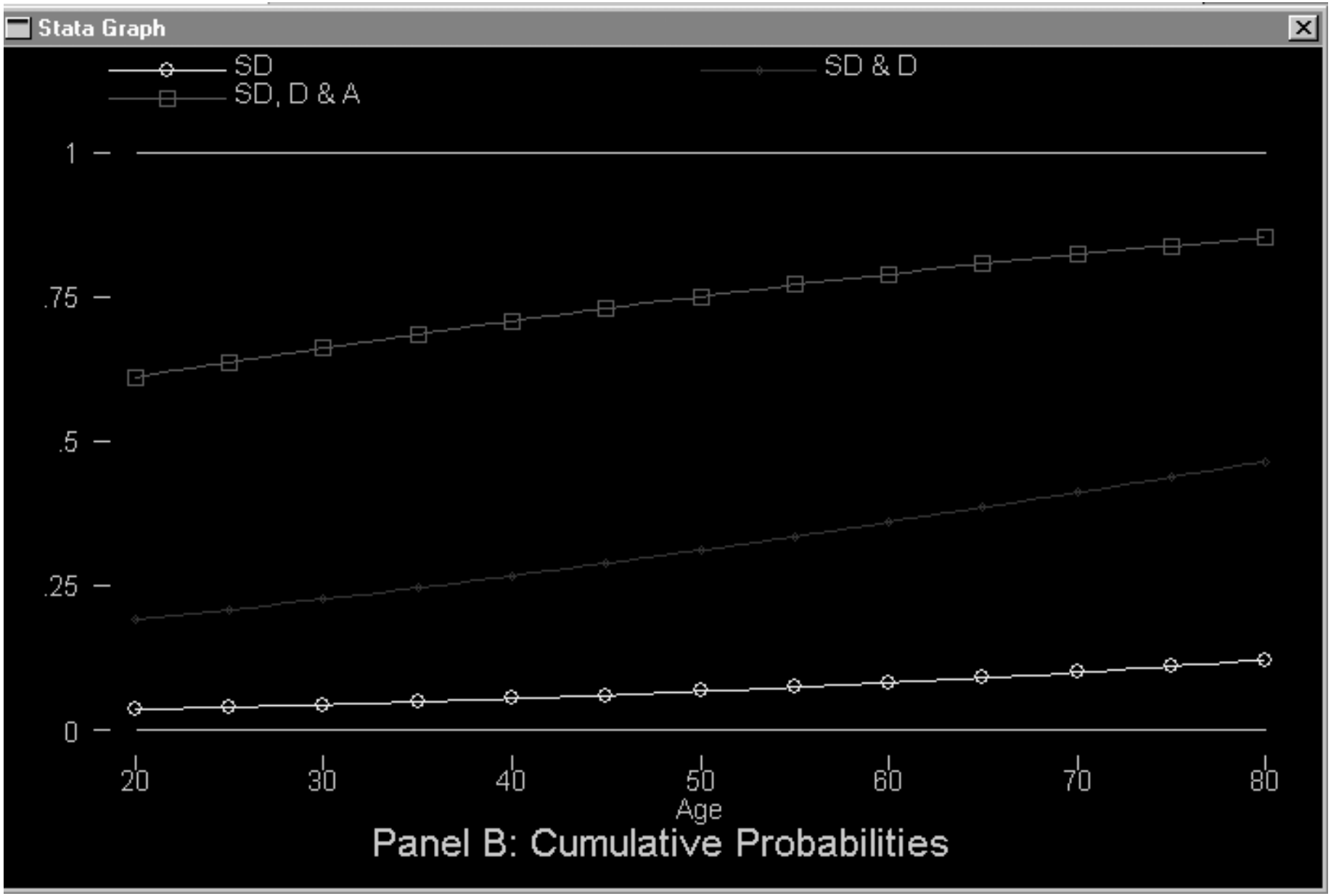


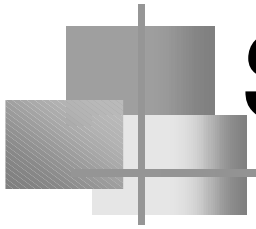
# SPost command: `prgen`

- Adds pseudovariables to data that can be used to generate plots of how predicted probability changes over range of continuous independent variable
- `from()` and `to()` options specify range of independent variable

Stata Results

```
. ologit warm yr89 male white age ed prst, nolog

Ordered logit estimates                           Number of obs   =       2293
                                                  LR chi2(6)      =     301.72
                                                  Prob > chi2     =     0.0000
Log likelihood = -2844.9123                       Pseudo R2       =     0.0504

------------------------------------------------------------------------------
        warm |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        yr89 |    .5239025   .0798988     6.56   0.000     .3673037    .6805013
        male |   -.7332997   .0784827    -9.34   0.000    -.8871229   -.5794766
       white |   -.3911595   .1183808    -3.30   0.001    -.6231815   -.1591374
         age |   -.0216655   .0024683    -8.78   0.000    -.0265032   -.0168278
          ed |    .0671728    .015975     4.20   0.000     .0358624    .0984831
        prst |    .0060727   .0032929     1.84   0.065    -.0003813    .0125267
-------------+----------------------------------------------------------------
       _cut1 |   -2.465362   .2389126          (Ancillary parameters)
       _cut2 |    -.630904   .2333155
       _cut3 |    1.261854   .2340179
------------------------------------------------------------------------------

. prgen age, from(20) to(80) generate(w89) x(male=0 yr89=1) ncases(13)

ologit: Predicted values as age varies from 20 to 80.

          yr89       male      white        age         ed       prst
x=           1          0   .8765809  44.935456  12.218055  39.585259

. label var w89s1 "SD"

. label var w89s2 "SD & D"

. label var w89s3 "SD, D & A"

. graph w89s1 w89s2 w89s3 w89x, /*
> */ title("Panel B: Cumulative Probabilities") b2("Age") /*
> */ xlabel(20,30,40,50,60,70,80) ylabel(0,.25,.50,.75,1.0) xscale(20,80) /*
> */ yscale(0,1) s(0dST) connect(sss) yline(0,1) gap(4) noaxis /*
> */ saving(tmp2.gph, replace)
(note: file tmp2.gph not found)
```


Stata Graph
SD
SD, D & A
SD & D
1
.75
.5
.25
0
20
30
40
50
60
70
80
Age
Panel B: Cumulative Probabilities



# SPost command: `praccum`

- More flexible (but harder to use) syntax that works with more complex model specifications

# SPost command: `prchange`

- Computes marginal and discrete change
- Discrete change from min->max, 0->1, as x increases by 1 unit, and as x increases by 1 sd

Stata Results

```
. logit lfp k5 k618 age wc hc lwg inc, nolog

Logit estimates                                   Number of obs   =        753
                                                  LR chi2(7)      =     124.48
                                                  Prob > chi2     =     0.0000
Log likelihood = -452.63296                       Pseudo R2       =     0.1209

------------------------------------------------------------------------------
         lfp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          k5 |  -1.462913   .1970006    -7.43   0.000    -1.849027   -1.076799
        k618 |  -.0645707   .0680008    -0.95   0.342    -.1978499    .0687085
         age |  -.0628706   .0127831    -4.92   0.000    -.0879249   -.0378162
          wc |   .8072738   .2299799     3.51   0.000     .3565215    1.258026
          hc |   .1117336   .2060397     0.54   0.588    -.2920969     .515564
         lwg |   .6046931   .1508176     4.01   0.000     .3090961    .9002901
         inc |  -.0344464   .0082084    -4.20   0.000    -.0505346   -.0183583
       _cons |    3.18214   .6443751     4.94   0.000     1.919188    4.445092
------------------------------------------------------------------------------

. prchange

logit: Changes in Predicted Probabilities for lfp

      min->max      0->1      -+1/2     -+sd/2  MargEfct
k5    -0.6361   -0.3499   -0.3428   -0.1849   -0.3569
k618  -0.1278   -0.0156   -0.0158   -0.0208   -0.0158
age   -0.4372   -0.0030   -0.0153   -0.1232   -0.0153
wc     0.1881    0.1881    0.1945    0.0884    0.1969
hc     0.0272    0.0272    0.0273    0.0138    0.0273
lwg    0.6624    0.1499    0.1465    0.0865    0.1475
inc   -0.6415   -0.0068   -0.0084   -0.0975   -0.0084

          NotInLF      inLF
Pr(y|x)    0.4222    0.5778

             k5        k618       age        wc         hc        lwg       inc
   x=  .237716   1.35325   42.5378   .281541   .391766   1.09711   20.129
sd(x)=  .523959   1.31987   8.07257   .450049   .488469   .587556   11.6348

.
```

# SPost command: `prcounts`

- Akin to `predict`, but generates new variables that contain the predicted probabilities of observing counts 0 through specified value (default=9)
- `prcounts dog` will generate `dogp0`, `dogp1` ... `dogp9`, as well as cumulative probabilities `dogs0` ... `dogs9`

# SPost command: `listcoef`

- `listcoef, std` – standardized coefficients (x-standardized, y-standardized, fully standardized)
- `listcoef, factor` – factor change in the odds/expected count
- `listcoef, percent` – percent change in the odds/expected count

Stata Results

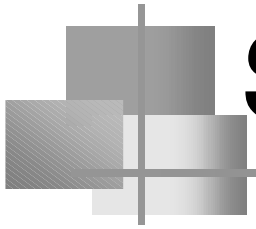
```
. listcoef, std

logit (N=753): Unstandardized and Standardized Estimates

 Observed SD: .49562951
   Latent SD: 2.0500391

  Odds of: inLF vs NotInLF
```

| lfp | b | z | P>\|z\| | bStdX | bStdY | bStdXY | SDofX |
|---|---|---|---|---|---|---|---|
| k5 | -1.46291 | -7.426 | 0.000 | -0.7665 | -0.7136 | -0.3739 | 0.5240 |
| k618 | -0.06457 | -0.950 | 0.342 | -0.0852 | -0.0315 | -0.0416 | 1.3199 |
| age | -0.06287 | -4.918 | 0.000 | -0.5075 | -0.0307 | -0.2476 | 8.0726 |
| wc | 0.80727 | 3.510 | 0.000 | 0.3633 | 0.3938 | 0.1772 | 0.4500 |
| hc | 0.11173 | 0.542 | 0.588 | 0.0546 | 0.0545 | 0.0266 | 0.4885 |
| lwg | 0.60469 | 4.009 | 0.000 | 0.3553 | 0.2950 | 0.1733 | 0.5876 |
| inc | -0.03445 | -4.196 | 0.000 | -0.4008 | -0.0168 | -0.1955 | 11.6348 |

```
. listcoef, percent

logit (N=753): Percentage Change in Odds

  Odds of: inLF vs NotInLF
```

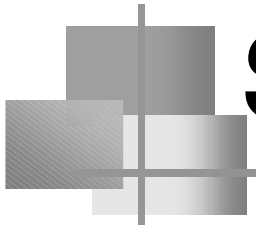
| lfp | b | z | P>\|z\| | % | %StdX | SDofX |
|---|---|---|---|---|---|---|
| k5 | -1.46291 | -7.426 | 0.000 | -76.8 | -53.5 | 0.5240 |
| k618 | -0.06457 | -0.950 | 0.342 | -6.3 | -8.2 | 1.3199 |
| age | -0.06287 | -4.918 | 0.000 | -6.1 | -39.8 | 8.0726 |
| wc | 0.80727 | 3.510 | 0.000 | 124.2 | 43.8 | 0.4500 |
| hc | 0.11173 | 0.542 | 0.588 | 11.8 | 5.6 | 0.4885 |
| lwg | 0.60469 | 4.009 | 0.000 | 83.1 | 42.7 | 0.5876 |
| inc | -0.03445 | -4.196 | 0.000 | -3.4 | -38.0 | 11.6348 |

```
.
```



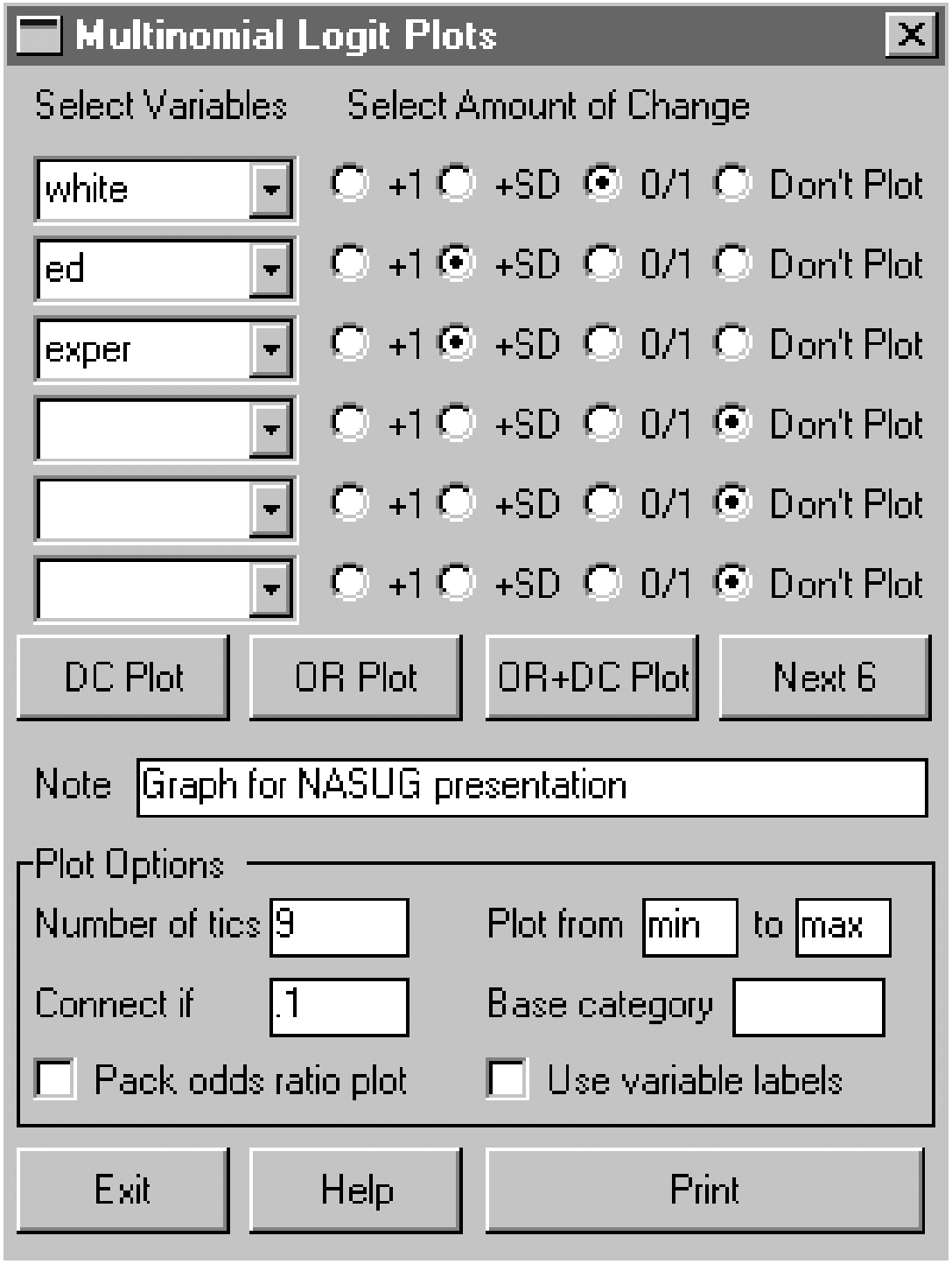
# SPost command: `mlogtest`

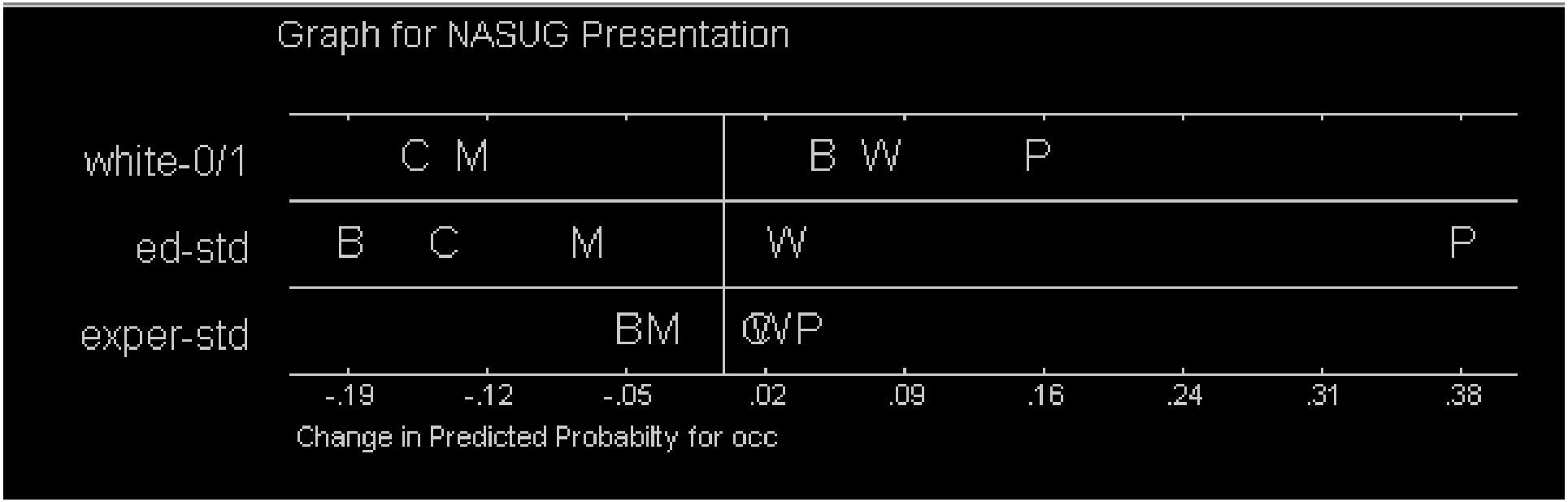
- Wald or LR test whether the effect of an independent variable is zero across all equations
- Wald or LR test whether a pair of outcomes is indistinguishable
- Hausman or Small-Hsiao tests of the IIA assumption

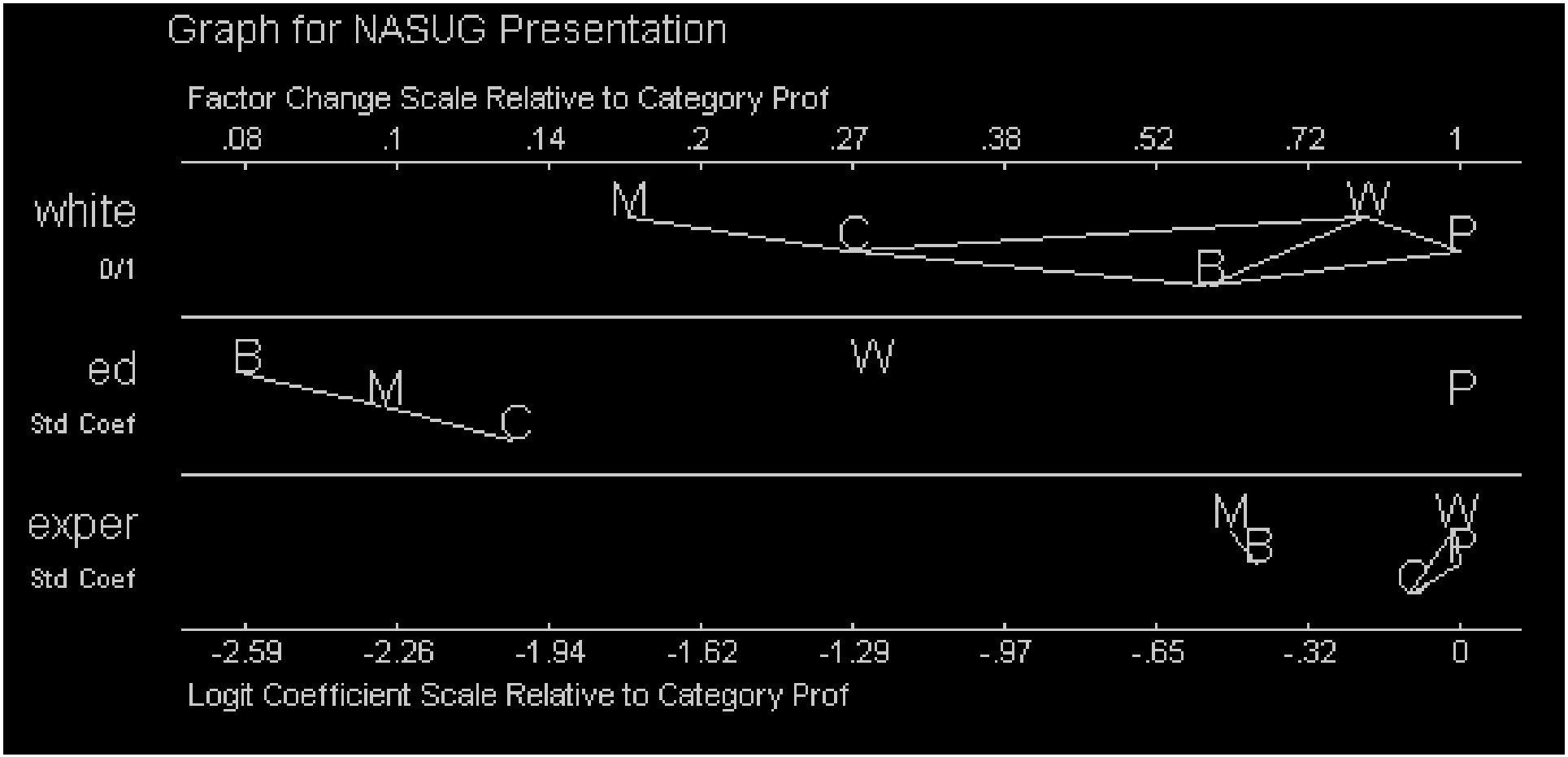


# SPost command: `mlogview`

- Dialog box interface for generating **discrete change** plots or **odds ratio** plots for `mlogit` models


Multinomial Logit Plots
Select Variables
Select Amount of Change
white
ed
exper
+1 +SD 0/1 Don't Plot
DC Plot
OR Plot
OR+DC Plot
Next 6
Note
Graph for NASUG presentation
Plot Options
Number of tics
9
Plot from
min
to
max
Connect if
.1
Base category
Pack odds ratio plot
Use variable labels
Exit
Help
Print


Graph for NASUG Presentation
white-0/1
ed-std
exper-std
C M
B W
P
B C M
W
P
BM
@VP
-.19
-.12
-.05
.02
.09
.16
.24
.31
.38
Change in Predicted Probabilty for occ


Graph for NASUG Presentation
Factor Change Scale Relative to Category Prof
.08
.1
.14
.2
.27
.38
.52
.72
1
white
0/1
M
C
B
W
P
ed
Std Coef
B
M
C
W
P
exper
Std Coef
M
B
W
P
-2.59
-2.26
-1.94
-1.62
-1.29
-.97
-.65
-.32
0
Logit Coefficient Scale Relative to Category Prof