Notes for laboratory session 10

Logistic regression analysis

Use the contraceptive use dataset and repeat the logistic regression analysis of the two-factor ("more", "age") additive model:

```
. char more[omit] 0

. xi: logit cuse i.more i.age [freq=N], nolog
i.more                Imore_0-1     (naturally coded; Imore_0 omitted)
i.age                 Iage_1-4      (naturally coded; Iage_1 omitted)

Logit estimates                                   Number of obs   =       1607
                                                  LR chi2(4)      =     128.88
                                                  Prob > chi2     =     0.0000
Log likelihood = -937.40449                       Pseudo R2       =     0.0643

------------------------------------------------------------------------------
        cuse |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
-------------+----------------------------------------------------------------
     Imore_1 |   -.824092   .1171128     -7.037   0.000      -1.053629   -.5945552
      Iage_2 |   .3678306   .1753673      2.097   0.036       .024117    .7115443
      Iage_3 |   .8077888   .1597533      5.056   0.000       .494678      1.1209
      Iage_4 |   1.022618   .2039337      5.014   0.000       .6229158    1.422321
       _cons |  -.8698414   .1571298     -5.536   0.000      -1.17781   -.5618727
------------------------------------------------------------------------------
```

Analysis as an ordinal logistic regression

Now use the ordinal logistic regression approach to fit the same model.

```
. xi: ologit cuse i.more i.age [freq=N], nolog table
i.more                Imore_0-1     (naturally coded; Imore_0 omitted)
i.age                 Iage_1-4      (naturally coded; Iage_1 omitted)

Ordered logit estimates                           Number of obs   =       1607
                                                  LR chi2(4)      =     128.88
                                                  Prob > chi2     =     0.0000
Log likelihood = -937.40449                       Pseudo R2       =     0.0643
------------------------------------------------------------------------------
        cuse |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
-------------+----------------------------------------------------------------
     Imore_1 |   -.824092   .1171128     -7.037   0.000      -1.053629   -.5945552
      Iage_2 |   .3678306   .1753673      2.097   0.036       .024117    .7115443
      Iage_3 |   .8077888   .1597533      5.056   0.000       .494678      1.1209
      Iage_4 |   1.022618   .2039337      5.014   0.000       .6229158    1.422321
-------------+----------------------------------------------------------------
       _cut1 |   .8698414   .1571298             (Ancillary parameter)
------------------------------------------------------------------------------

        cuse |      Probability            Observed
-------------|-----------------------------------
          No |   Pr(      xb+u<_cut1)         0.6845
         Yes |   Pr(_cut1<xb+u)               0.3155
```

a) Compare the coefficients produced by the "logit" and the "ologit" STATA command.

Compare the results given by the table option in the logit command output with the following table.

```
. tab cuse [freq=N]

Contracepti |
     ve use |
   (Yes/No) |      Freq.     Percent        Cum.
------------+-----------------------------------
         No |       1100       68.45       68.45
        Yes |        507       31.55      100.00
------------+-----------------------------------
      Total |       1607      100.00
```

Predict the probabilities of contraceptive use and non-use via STATA predict command.

```
. predict p0 p1
(option p assumed; predicted probabilities)
```

b) Use hand calculations in order to obtain the predicted probability of contraceptive use among women 25-29 years old desiring more children. Verify your result using the listing below.

```
.  list age educat more cuse  p0 p1 if age==2 & more==1

           age     educat       more       cuse         p0         p1
  9.     25-29       High        Yes        Yes   .7901953   .2098047
 10.     25-29        Low        Yes         No   .7901953   .2098047
 11.     25-29       High        Yes         No   .7901953   .2098047
 16.     25-29        Low        Yes        Yes   .7901953   .2098047
```

Probit regression analysis

Now perform the same analysis using probit regression

```
. xi: probit cuse i.age i.more [freq=N], nolog
i.age                 Iage_1-4     (naturally coded; Iage_1 omitted)
i.more                Imore_0-1    (naturally coded; Imore_0 omitted)

Probit estimates                                  Number of obs   =       1607
                                                  LR chi2(4)      =     127.51
                                                  Prob > chi2     =     0.0000
Log likelihood = -938.09112                       Pseudo R2       =     0.0636

------------------------------------------------------------------------------
    cuse |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  Iage_2 |   .2086109   .1003457      2.079   0.038       .0119369     .405285
  Iage_3 |   .4685637   .0928326      5.047   0.000       .2866152    .6505122
  Iage_4 |   .6048679   .1226446      4.932   0.000       .3644889    .8452469
 Imore_1 |  -.4964618   .0714451     -6.949   0.000      -.6364916   -.3564319
   _cons |   -.515345   .0922618     -5.586   0.000      -.6961748   -.3345152
```

c) Compare the results of the probit and the logit model. Recall however that the logit coefficients are not standardized but must be divided $\pi/\sqrt{3}$.

Ordinal regression

Use the tumour data set and fit a bivariate (sex , therapy) ordinal regression model for the outcome variable "outc".

```
. char sex[omit] 2

. xi: ologit outc i.sex i.therapy, nolog tab
i.sex                 Isex_1-2     (naturally coded; Isex_2 omitted)
i.therapy             Ithera_0-1   (naturally coded; Ithera_0 omitted)

Ordered logit estimates                           Number of obs   =        299
                                                  LR chi2(2)      =      10.91
                                                  Prob > chi2     =     0.0043
Log likelihood = -394.52832                       Pseudo R2       =     0.0136

------------------------------------------------------------------------------
    outc |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  Isex_1 |   .5413938   .2871816      1.885   0.059      -.0214717    1.104259
Ithera_1 |   -.580685   .2121478     -2.737   0.006      -.9964871    -.164883
---------+--------------------------------------------------------------------
   _cut1 |  -.7766492   .2880856             (Ancillary parameters)
   _cut2 |   .7906273   .2866223
   _cut3 |    1.84145   .3056123
------------------------------------------------------------------------------

    outc |      Probability            Observed
---------|-----------------------------------
Progress |   Pr(      xb+u<_cut1)         0.2843
No chang |   Pr(_cut1<xb+u<_cut2)         0.3612
Partial  |   Pr(_cut2<xb+u<_cut3)         0.1906
Complete |   Pr(_cut3<xb+u)               0.1639
```

d) What is the interpretation of the model coefficients?

Probit analysis

Analyze the previous example using probit analysis.

```
. xi: oprobit outc i.sex i.therapy, nolog
i.sex                 Isex_1-2     (naturally coded; Isex_2 omitted)
i.therapy             Ithera_0-1   (naturally coded; Ithera_0 omitted)

Ordered probit estimates                          Number of obs   =        299
                                                  LR chi2(2)      =      10.79
                                                  Prob > chi2     =     0.0045
Log likelihood =  -394.5871                       Pseudo R2       =     0.0135

------------------------------------------------------------------------------
    outc |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  Isex_1 |   .3401406    .174902      1.945   0.052       -.002661    .6829422
Ithera_1 |  -.3344764    .125435     -2.667   0.008      -.5803245   -.0886282
---------+--------------------------------------------------------------------
   _cut1 |   -.459358    .176613             (Ancillary parameters)
   _cut2 |   .5050695   .1760197
   _cut3 |   1.122025   .1836877
```

e) Compare the results of the previous two approaches.