Notes for laboratory session 9

Produce two 2x2 tables corresponding to the association between "Mammograph experience" and "Family history". In the first table use only the "Within one year" and "Never" categories while in the second use the "Over a year ago" and "Never" categories of the "Mammograph experience" variable.

```
. tab ME hist if ME==0 | ME==1

     Mammograph |     Fam. history
     experience |        No        Yes |     Total
----------------+----------------------+----------
          Never |       220         14 |       234
Within one year |        85         19 |       104
----------------+----------------------+----------
          Total |       305         33 |       338

. tab ME hist if ME==0 | ME==2

     Mammograph |     Fam. history
     experience |        No        Yes |     Total
----------------+----------------------+----------
          Never |       220         14 |       234
Over a year ago |        63         11 |        74
----------------+----------------------+----------
          Total |       283         25 |       308
```

a) *Calculate the Odds Ratios associated with the tables above.*

Now fit the logistic regression models corresponding to the previous tables .

```
. xi: logit ME i.hist if ME==1 | ME==0, nolog
i.hist            Ihist_0-1    (naturally coded; Ihist_0 omitted)

Logit estimates                                   Number of obs   =        338
                                                  LR chi2(1)      =      11.32
                                                  Prob > chi2     =     0.0008
Log likelihood = -202.96528                       Pseudo R2       =     0.0271

------------------------------------------------------------------------------
      ME |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
 Ihist_1 |   1.256358   .3746603      3.353   0.001       .5220373    1.990679
   _cons |  -.9509763   .1277112     -7.446   0.000      -1.201286   -.7006669
------------------------------------------------------------------------------
. xi: logit ME i.hist if ME==2 | ME==0, nolog
i.hist            Ihist_0-1    (naturally coded; Ihist_0 omitted)

Logit estimates                                   Number of obs   =        308
                                                  LR chi2(1)      =       5.26
                                                  Prob > chi2     =     0.0218
Log likelihood = -167.19417                       Pseudo R2       =     0.0155

------------------------------------------------------------------------------
      ME |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
 Ihist_1 |   1.009331   .4274999      2.361   0.018       .1714464    1.847215
   _cons |  -1.250493   .1428932     -8.751   0.000      -1.530558   -.9704273
------------------------------------------------------------------------------
```

Now perform the above two logistic regressions in one step using the multinomial logit (`mlogit`) command of STATA :

```
. xi: mlogit ME i.hist, nolog
i.hist            Ihist_0-1    (naturally coded; Ihist_0 omitted)

Multinomial regression                         Number of obs   =        412
                                               LR chi2(2)      =      12.86
                                               Prob > chi2     =     0.0016
Log likelihood = -396.16997                    Pseudo R2       =     0.0160
------------------------------------------------------------------------------
      ME |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
Within o |
 Ihist_1 |   1.256358   .3746603      3.353   0.001       .5220372    1.990679
   _cons |  -.9509763   .1277112     -7.446   0.000      -1.201286   -.7006669
---------+--------------------------------------------------------------------
Over a y |
 Ihist_1 |   1.009331   .4274998      2.361   0.018       .1714466    1.847215
   _cons |  -1.250493   .1428932     -8.751   0.000      -1.530558   -.9704273
------------------------------------------------------------------------------
(Outcome ME==Never is the comparison group)

. est store A1
```

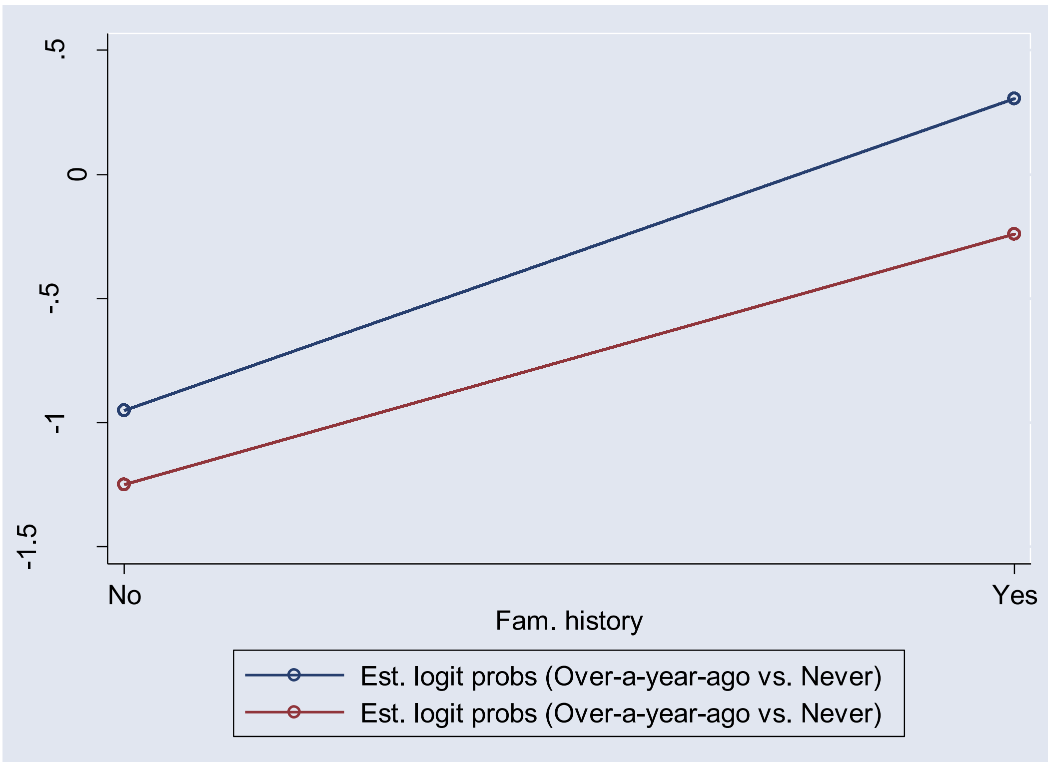
b) *Compare the output of the “mlogit” command to the output of the two “logit” commands.*

c) *What is the interpretation of the coefficients in the “mlogit” command output. Compare whith the results of question a).*

Graphical inspection of the results

We can inspect graphically the results as follows:

```
. quietly xi: mlogit ME i.hist
. predict p1 p2 p3
(option p assumed; predicted probabilities)
. gen logit_1=log(p2/p1)
. label var logit_1 "Est. logit probs (Within-one-year vs. Never)"
. gen logit_2=log(p3/p1)
. label var logit_2 "Est. logit probs (Over-a-year-ago vs. Never)"
. sc logit_1 logit_2 hist, c(l l)  xlab(0 1, val) legend(row(2)) ylab()
yti("Estimated logits")
```



Testing the equality of the two odds ratios

Test whether OR_1=OR_2 by testing whether the odds ratio that corresponds to the following table is one:

```
. tab ME hist if ME==1 | ME==2,chi

     Mammograph |     Fam. history
     experience |        No        Yes |     Total
----------------+----------------------+----------
Within one year |        85         19 |       104
Over a year ago |        63         11 |        74
----------------+----------------------+----------
          Total |       148         30 |       178

        Pearson chi2(1) =   0.3576   Pr = 0.550
```

d) *Calculate the OR and its SE in order to produce the relevant statistic and test whether OR=1.*

We can use this alternative approach to test the same hypothesis (OR_1=OR_2) :

```
. test [1]

 ( 1)  [Within o]Ihist_1 = 0.0

           chi2(  1) =   11.24
         Prob > chi2 =    0.0008

. test [2]

 ( 1)  [Over a y]Ihist_1 = 0.0

           chi2(  1) =    5.57
         Prob > chi2 =    0.0182

. test [1=2]

 ( 1)  [Within o]Ihist_1 - [Over a y]Ihist_1 = 0.0

           chi2(  1) =    0.36
         Prob > chi2 =    0.5505
```

e) *Notice the relation between the chi-square statistics in the first two "`test`" commands and the z statistics in the previous "`mlogit`" command.*

Now fit the null model in order to check the significance of "`family history`" by using a likelihood-ratio test approach.

```
. mlogit ME, nolog

Multinomial regression                            Number of obs   =        412
                                                  LR chi2(0)      =       0.00
                                                  Prob > chi2     =          .
Log likelihood = -402.59901                       Pseudo R2       =     0.0000

------------------------------------------------------------------------------
      ME |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
Within o |
   _cons |  -.8109302   .1178511     -6.881   0.000      -1.041914   -.5799462
---------+--------------------------------------------------------------------
Over a y |
   _cons |  -1.151256    .133368     -8.632   0.000      -1.412652   -.8898596
------------------------------------------------------------------------------
(Outcome ME==Never is the comparison group)

. est store A0
```

```
. lrtest A0 A1

likelihood-ratio test                                   LR chi2(2)  =     12.86
(Assumption: A0 nested in A1)                           Prob > chi2 =    0.0016
```

f) Calculate the likelihood-ratio test manually.

The significance of "`family history`" can also be checked by a Wald type test:

```
. quietly xi: mlogit ME i.hist, nolog

. test Ihist_1

 ( 1)  [Within o]Ihist_1 = 0.0
 ( 2)  [Over a y]Ihist_1 = 0.0

           chi2(  2) =   12.01
         Prob > chi2 =    0.0025
```

g) What is the conclusion of the above tests?

Incorporating a polytomous covariate

Tabulate the data according to the variables "`mammograph experience`" and "`detc`"

```
. tab ME detc

      Mammograph |        Likely find cancer
      experience | Not likel  Somewhat   Very like |     Total
-----------------+---------------------------------+----------
           Never |        13         77        144 |       234
 Within one year |         1         12         91 |       104
 Over a year ago |         4         16         54 |        74
-----------------+---------------------------------+----------
           Total |        18        105        289 |       412
```

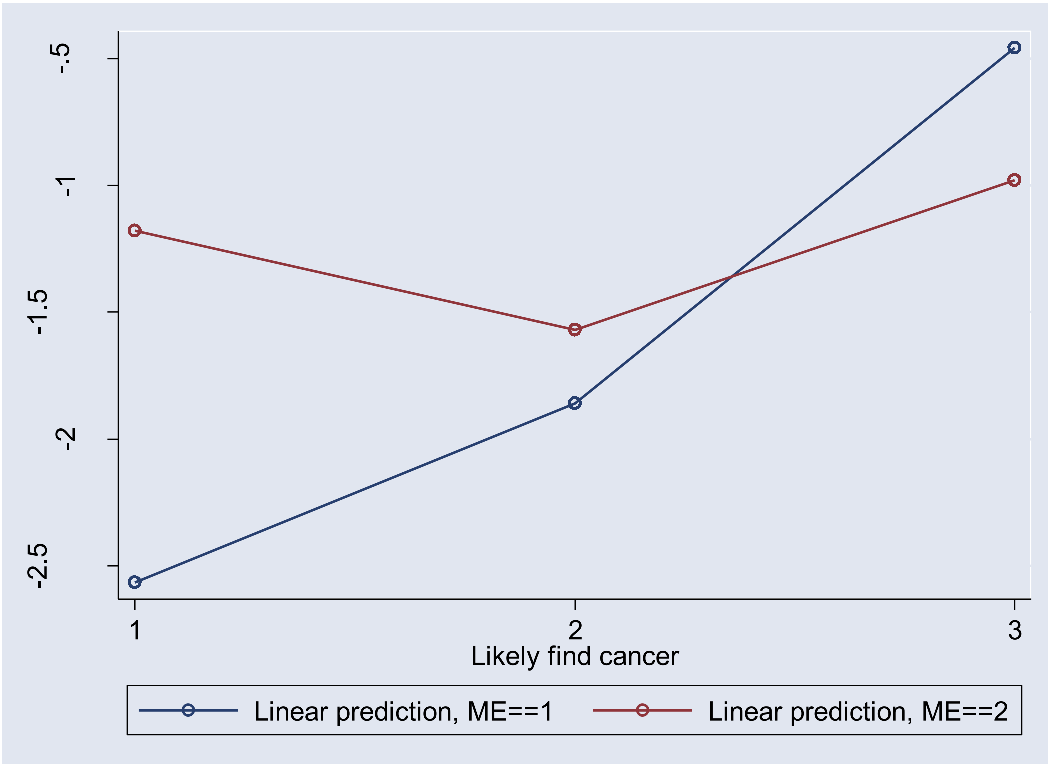
Now define the level 1 of the factor "`detc`" as reference category and then use the "`mlogit`" STATA command in order to investigate the association between these two variables.

```
. char detc[omit] 1
. xi: mlogit ME i.detc, nolog
i.detc              Idetc_1-3    (naturally coded; Idetc_1 omitted)

Multinomial regression                            Number of obs   =        412
                                                  LR chi2(4)      =      26.80
                                                  Prob > chi2     =     0.0000
Log likelihood = -389.20054                       Pseudo R2       =     0.0333

------------------------------------------------------------------------------
      ME |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
Within o |
 Idetc_2 |   .7060506   1.083136      0.652   0.514      -1.416856    2.828958
 Idetc_3 |   2.105996   1.046325      2.013   0.044       .0552361    4.156755
   _cons |  -2.564949    1.03772     -2.472   0.013      -4.598843   -.5310556
---------+--------------------------------------------------------------------
Over a y |
 Idetc_2 |  -.3925617   .6343589     -0.619   0.536      -1.635882     .850759
 Idetc_3 |   .1978257   .5936221      0.333   0.739      -.9656522    1.361304
   _cons |  -1.178655   .5717729     -2.061   0.039      -2.299309   -.0580007
------------------------------------------------------------------------------
(Outcome ME==Never is the comparison group)
```

h) Try to interpret the coefficients of the `Idetc_2` and `Idetc_3` dummy variables in terms of ORs produced by appropriate sub-tables of the 3x3 table above.

Graphical inspection of the model

```
sort detc
cap drop logit_1 logit_2
predict logit_1,xb outcome(1)
predict logit_2,xb outcome(2)
sc logit_1 logit_2 detc, c(l l)  xlab(1 2 3) ylab() l1("log odds")
```



We can now use the "`test`" command for hypotheses testing

```
. test [1]

 ( 1)  [Within o]Idetc_2 = 0.0
 ( 2)  [Within o]Idetc_3 = 0.0

           chi2(  2) =   20.41
         Prob > chi2 =    0.0000

. test [2]

 ( 1)  [Over a y]Idetc_2 = 0.0
 ( 2)  [Over a y]Idetc_3 = 0.0

           chi2(  2) =    3.46
         Prob > chi2 =    0.1773

. test [1=2]

 ( 1)  [Within o]Idetc_2 - [Over a y]Idetc_2 = 0.0
 ( 2)  [Within o]Idetc_3 - [Over a y]Idetc_3 = 0.0

           chi2(  2) =    6.20
         Prob > chi2 =    0.0450
```

i) *State the null hypotheses for the three "`test`" commands.*

Assessment of the significance of a continuous factor

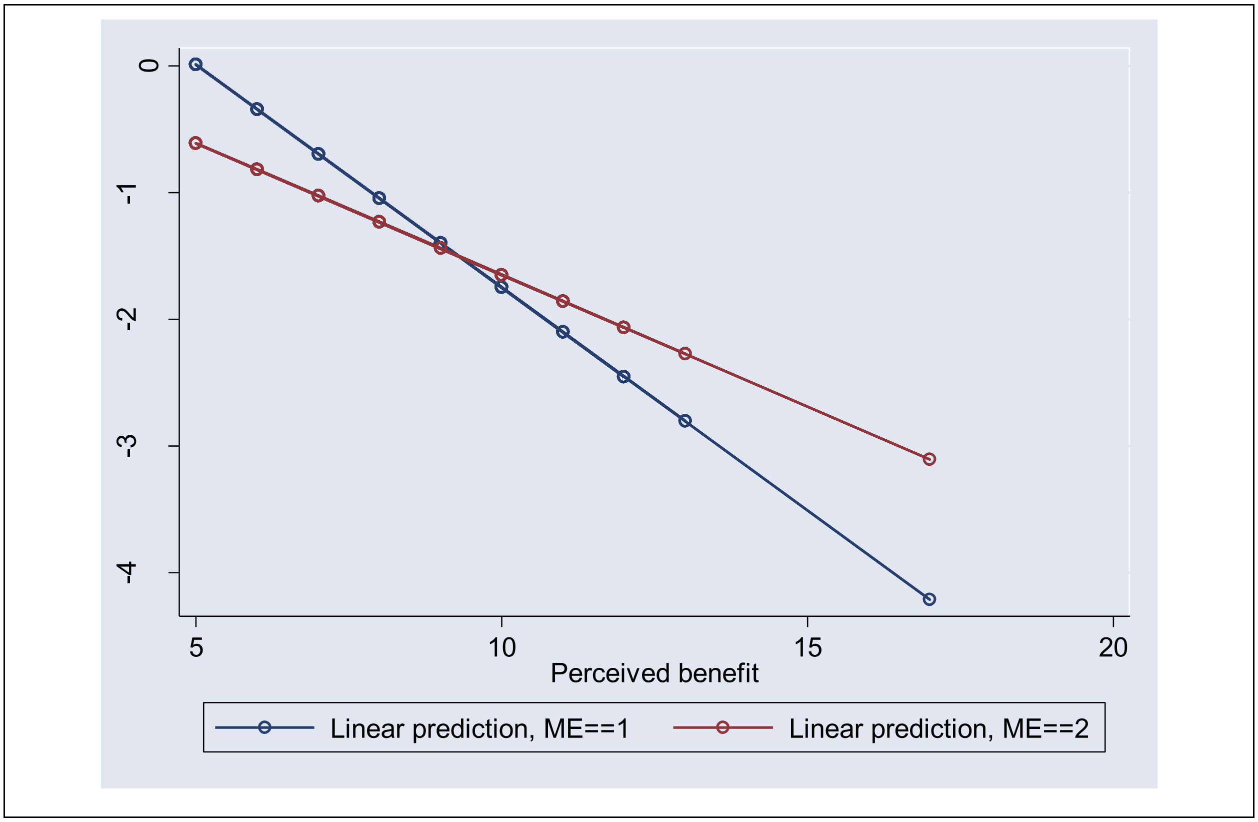
Fit the multinomial logistic regression model for the effect of perceived benefit (pb) of mammography (higher scores denote a smaller perceived benefit).

```
. mlogit ME pb,nolog

Multinomial regression                            Number of obs   =        412
                                                  LR chi2(2)      =      35.25
                                                  Prob > chi2     =     0.0000
Log likelihood = -384.97236                       Pseudo R2       =     0.0438

------------------------------------------------------------------------------
      ME |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
Within o |
      pb |   -.351685   .0666009     -5.280   0.000      -.4822205   -.2211496
   _cons |    1.76865    .484798      3.648   0.000       .8184631    2.718836
---------+--------------------------------------------------------------------
Over a y |
      pb |   -.207975   .0684675     -3.038   0.002      -.3421688   -.0737812
   _cons |   .4314007   .5228924      0.825   0.409      -.5934496    1.456251
------------------------------------------------------------------------------
(Outcome ME==Never is the comparison group)
```

k) What is the interpretation of the "pb" coefficients?

We can see graphically the model as follows:

```
. predict lhat1, xb outcome(1)

. predict lhat2, xb outcome(2)

. sc lhat* pb, xlab() ylab() c(l l)
```


0
-1
-2
-3
-4
5
10
15
20
Perceived benefit
Linear prediction, ME==1
Linear prediction, ME==2

Testing the effect of a continuous covariate

The overall effect of the perceived benefit on the likelihood of a mammogram is tested as follows:

```
. test pb

 ( 1)  [Within o]pb = 0.0
 ( 2)  [Over a y]pb = 0.0

           chi2(  2) =   31.00
         Prob > chi2 =    0.0000
```

On the other hand, we can test whether the relationship between pb and ME is the same for both levels 1 and 2 compared to level 0 :

```
. test [1]pb=[2]pb

 ( 1)  [Within o]pb - [Over a y]pb = 0.0

           chi2(  1) =    3.02
         Prob > chi2 =    0.0821
```

l) Interpret the result of the previous test command and compare it with the graphical representation of the model.

The method of Begg & Gray (Biometrika, 1984)

Now use the method of Begg & Gray by fitting two logistic regression models instead of one multinomial logistic regression model

```
. xi: logit ME i.sympt pb i.hist BSE i.detc if ME==1|ME==0, nolog

Log likelihood = -160.14267                       Pseudo R2       =     0.2324

------------------------------------------------------------------------------
      ME |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
Isympt_2 |   .2333352   .9354078      0.249   0.803       -1.60003    2.066701
Isympt_3 |   2.003157   .7897107      2.537   0.011       .4553526    3.550961
Isympt_4 |   2.526276   .7863638      3.213   0.001       .9850313    4.067521
      pb |  -.2127205   .0767566     -2.771   0.006      -.3631606   -.0622804
 Ihist_1 |   1.415273   .4687315      3.019   0.003       .4965762     2.33397
     BSE |   1.400521   .5382657      2.602   0.009         .34554    2.455503
 Idetc_2 |     .20745   1.166733      0.178   0.859      -2.079304    2.494204
 Idetc_3 |   1.050927   1.126617      0.933   0.351      -1.157202    3.259057
   _cons |  -3.376979   1.540057     -2.193   0.028      -6.395435   -.3585223
------------------------------------------------------------------------------
```

```
. xi: logit ME i.sympt pb i.hist BSE i.detc if ME==2|ME==0, nolog

Log likelihood = -152.15547                       Pseudo R2       =     0.1040

------------------------------------------------------------------------------
      ME |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
Isympt_2 |  -.2744739   .6472962     -0.424   0.672      -1.543151    .9942034
Isympt_3 |   .8085675   .5421184      1.491   0.136      -.2539651      1.8711
Isympt_4 |   1.209314   .5479505      2.207   0.027        .135351    2.283278
      pb |  -.1426345    .075692     -1.884   0.060      -.2909881    .0057191
 Ihist_1 |   1.156508   .4735987      2.442   0.015       .2282712    2.084744
     BSE |   1.018466   .5157325      1.975   0.048       .0076488    2.029283
 Idetc_2 |   -.778484   .7179777     -1.084   0.278      -2.185694    .6287264
 Idetc_3 |   -.566554    .686941     -0.825   0.410      -1.912934    .7798256
   _cons |  -1.156001   1.112774     -1.039   0.299      -3.336999    1.024996
```

The full multinomial logistic regression is as follows:

```
.  xi: mlogit ME i.sympt pb i.hist BSE i.detc , nolog
i.sympt              Isympt_1-4    (naturally coded; Isympt_1 omitted)
i.hist               Ihist_0-1     (naturally coded; Ihist_0 omitted)
i.detc               Idetc_1-3     (naturally coded; Idetc_1 omitted)

Multinomial regression                            Number of obs   =        412
                                                  LR chi2(16)     =     111.30
                                                  Prob > chi2     =     0.0000
Log likelihood = -346.95096                       Pseudo R2       =     0.1382

------------------------------------------------------------------------------
      ME |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
Within o |
Isympt_2 |   .1100372   .9227608      0.119   0.905      -1.698541    1.918615
Isympt_3 |   1.924708   .7775975      2.475   0.013       .4006449    3.448771
Isympt_4 |   2.456993   .7753323      3.169   0.002       .9373693    3.976616
      pb |  -.2194368   .0755139     -2.906   0.004      -.3674414   -.0714323
 Ihist_1 |   1.366239   .4375196      3.123   0.002       .5087163    2.223762
     BSE |   1.291666    .529891      2.438   0.015       .2530992    2.330234
 Idetc_2 |   .0170207   1.161896      0.015   0.988      -2.260254    2.294296
 Idetc_3 |   .9041379   1.126822      0.802   0.422      -1.304393    3.112668
   _cons |   -2.99875    1.53922     -1.948   0.051      -6.015566    .0180663
---------+--------------------------------------------------------------------
Over a y |
Isympt_2 |  -.2900833   .6440636     -0.450   0.652      -1.552425    .9722582
Isympt_3 |   .8173136   .5397922      1.514   0.130      -.2406596    1.875287
Isympt_4 |   1.132239   .5476704      2.067   0.039       .0588252    2.205654
      pb |  -.1482068   .0763686     -1.941   0.052      -.2978866    .0014729
 Ihist_1 |   1.065436    .459396      2.319   0.020       .1650366    1.965836
     BSE |   1.052144   .5149894      2.043   0.041       .0427838    2.061505
 Idetc_2 |  -.9243928   .7137382     -1.295   0.195      -2.323294    .4745083
 Idetc_3 |  -.6905329   .6871078     -1.005   0.315      -2.037239    .6561736
   _cons |  -.9860915   1.111832     -0.887   0.375      -3.165242    1.193059
------------------------------------------------------------------------------
(Outcome ME==Never is the comparison group)
```

m) Compare the results of the above two approaches.