GENERALIZED LINEAR MODELS
Multinomial Logistic Regression

Γιώτα Τουλούμη

Καθηγήτρια Βιοστατιστικής και Επιδημιολογίας

Εργ. Υγιεινής, Επιδημιολογίας και Ιατρικής Στατιστικής

Ιατρική Σχολή Πανεπιστημίου Αθήνας

gtouloum@med.uoa.gr

Mammography Experience Study (Hosmer & Lemeshow, 2000)

Variable	Description	Codes/Values	Name
1	Identification Code	1-412	OBS
2	Mammography Experience	0 = Never 1 = Within One Year 2 = Over One Year Ago	ME
3	"You do not need a mamogram unless you develop symptoms"	1 = Strongly Agree 2 = Agree 3 = Disagree 4 = Strongly Disagree	SYMPT
4	Perveived benefit of mammography*	5 - 20	PB
5	Mother or Sister with a history of breast cancer	0 = No, 1 = Yes	HIST
6	"Has anyone taught you how to examine your own breasts: that is BSE"	0 = No, 1 = Yes	BSE
7	"How likely is it that a mamogram could find a new case of breast cancer"	1= Not likely 2 = Somewhat likely 3 = Very likely	DETC

*The variable PB is the sum of five scaled responses, each on a four point scale.
A low value is indicative of a woman with strong agreement with the benefits of mammography.

Multinomial logistic regression

Since the outcome variable ME takes on values 0, 1 or 2, we are involved in a multinomial (polytomous) logistic regression situation. We define the following models:

$$g_1(\mathbf{x}) = \log\left[\frac{P(Y=1|\mathbf{x})}{P(Y=0|\mathbf{x})}\right] = \beta_{1o} + \beta_{11}x_1 + \cdots + \beta_{1p}x_p = \mathbf{x}'\boldsymbol{\beta}_1 \quad \text{and} \quad g_2(\mathbf{x}) = \log\left[\frac{P(Y=2|\mathbf{x})}{P(Y=0|\mathbf{x})}\right] = \beta_{2o} + \beta_{21}x_1 + \cdots + \beta_{2p}x_p = \mathbf{x}'\boldsymbol{\beta}_2$$

Relative Risk

The probabilities for each outcome category are $\pi_0(\mathbf{x}) = P(Y=0|\mathbf{x}) = \dfrac{1}{1+e^{g_1(\mathbf{x})}+e^{g_2(\mathbf{x})}}$,

$\pi_1(\mathbf{x}) = P(Y=1|\mathbf{x}) = \dfrac{e^{g_1(\mathbf{x})}}{1+e^{g_1(\mathbf{x})}+e^{g_2(\mathbf{x})}}$ and $\pi_2(\mathbf{x}) = P(Y=2|\mathbf{x}) = \dfrac{e^{g_2(\mathbf{x})}}{1+e^{g_1(\mathbf{x})}+e^{g_2(\mathbf{x})}}$ or, in general,

(Hosmer & Lemeshow, 2000) $\pi_j(\mathbf{x}) = P(Y=j|\mathbf{x}) = \dfrac{e^{g_j(\mathbf{x})}}{\sum_{k=0}^{J} e^{g_k(\mathbf{x})}}$ with $g_o(\mathbf{x}) = 0$.

The multinomial logistic regression likelihood

If we consider three indicator variables Y_0, Y_1 and Y_2 such that $Y_j = 1$ if $Y = j$, j=0,1,2, then the multinomial logistic likelihood can be written (Hosmer & Lemeshow, 2000)

$$L(\boldsymbol{\pi},\boldsymbol{\beta}) = \prod_{i=1}^{n} \left[\pi_0(\mathbf{x}_i)^{y_{0i}} \pi_1(\mathbf{x}_i)^{y_{1i}} \pi_2(\mathbf{x}_i)^{y_{2i}} \right]$$

where $\boldsymbol{\beta}' = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')$ the coefficients corresponding to `ME==1` and `ME==2` respectively. The log-likelihood (realizing that $Y_0 = 1 - Y_1 - Y_2$) is:

$$l(\boldsymbol{\pi},\boldsymbol{\beta}) = \sum \left[y_{1i} \log(\pi_{1i}) + y_{2i} \log(\pi_{2i}) + (1 - y_{1i} - y_{2i}) \log(\pi_{0i}) \right]$$

$$= \sum \left[y_{1i} g_1(\mathbf{x}) + y_{2i} g_2(\mathbf{x}) - \log\left(1 + e^{g_1(\mathbf{x})} + e^{g_2(\mathbf{x})}\right) \right]$$

which is maximized in order to obtain the ML estimates of $\boldsymbol{\beta}$.

The two 2×2 tables corresponding to the logistic-regressions mentioned above are:

“Within one year” versus “Never”

```
. 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
```

“Over a year ago” versus “Never”

```
. 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
```

```
. 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
------------------------------------------------------------------------------
```

We can perform the above two logistic regressions in one step using the multinomial logit (`mlogit`) command of STATA is as follows:

```
. 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)

. lrtest, saving(1)
```

Notice that the mammography experience category "Never" has been used as the reference outcome

Comments: Comparing the output above to the two 2×2 tables shown earlier we see the following: The estimates of the coefficients and their interpretations are as follows:

1. $e^{\hat{\beta}_1} = e^{1.26} = \frac{(220)(19)}{(85)(14)} = 3.51$ is the estimate of the odds ratio referring to the first 2×2 table.

   Women with family history of breast cancer are 3.5 times more likely to have had a mammogram in the last year compared (versus not ever having had one) to women with no family history of breast cancer. The Wald test of significance for $\hat{\beta}_1$ is 3.353, which compared to a normal distribution results in a significant p value,

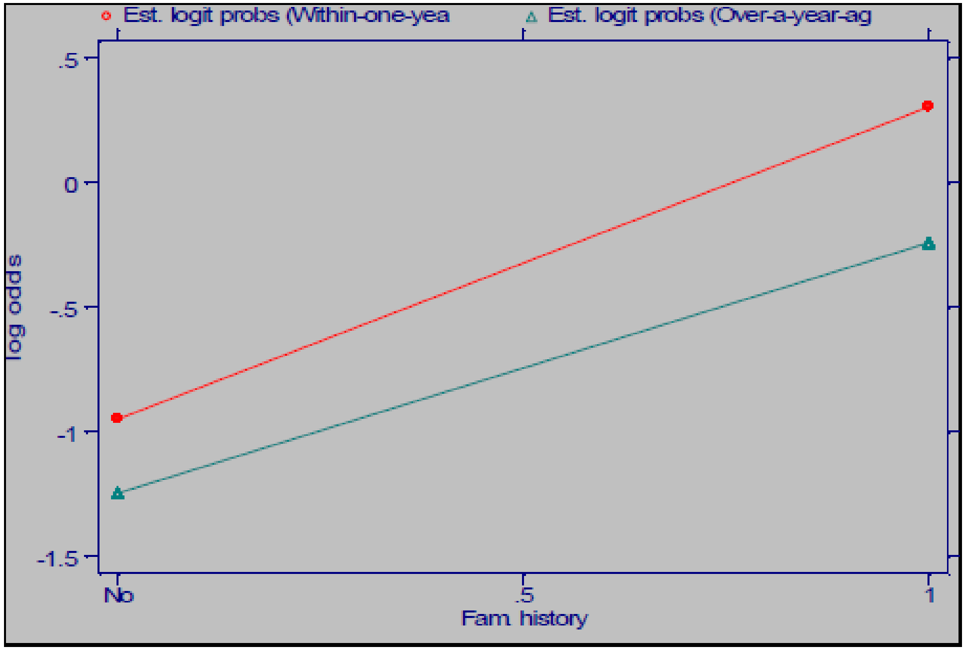
2. $e^{\hat{\beta}_2} = e^{1.01} = \frac{(220)(11)}{(63)(14)} = 2.74$ is the estimate of the odds ratio referring to the second 2×2 table.

   Women with a family history of breast cancer are 2.7 times more likely to have had a mammogram over a year ago (versus never having one) compared to women with no family history of breast cancer. The test of significance for $\hat{\beta}_2$ is 2.361, which compared to a normal distribution has a significant p value. This p value is close to the Pearson chi-square test associated with the second table.

Graphical inspection of the results

We can inspect graphically the results as follows:

```
. quietly xi: mlogit ME i.hist
. predict p0 p1 p2
(option p assumed; predicted probabilities)
. gen logit_1=log(p1/p0)
. label var logit_1 "Est. logit probs (Within-one-year vs. Never)"
. gen logit_2=log(p2/p0)
. label var logit_1 "Est. logit probs (Over-a-year-ago vs. Never)"
. graph logit_1 logit_2 hist, c(ll)  xlab(0 1) ylab border
```



Comments

Note p0 $=\dfrac{1}{1+e^{\hat\beta_{10}+\hat\beta_{11}}+e^{\hat\beta_{20}+\hat\beta_{21}}}=\dfrac{1}{1+e^{(-0.951)+1.256}+e^{(-1.250)+1.009}}=0.318=P(Y=0\,|\,X=1)=\hat\pi_0(1)$ and

Also, p0 $=\dfrac{1}{1+e^{\hat\beta_{10}}+e^{\hat\beta_{20}}}=\dfrac{1}{1+e^{(-0.951)}+e^{(-1.250)}}=0.598=P(Y=0\,|\,X=0)=\hat\pi_0(0)$ where X is `hist`.

Similarly, p1 $=\dfrac{e^{\hat\beta_{10}+\hat\beta_{11}}}{1+e^{\hat\beta_{10}+\hat\beta_{11}}+e^{\hat\beta_{20}+\hat\beta_{21}}}=\dfrac{e^{(-0.951)+1.256}}{1+e^{(-0.951)+1.256}+e^{(-1.250)+1.009}}=0.432=P(Y=1\,|\,X=1)=\hat\pi_1(1)$

while p1 $=\dfrac{e^{\hat\beta_{10}}}{1+e^{\hat\beta_{10}}+e^{\hat\beta_{20}}}=\dfrac{e^{(-0.951)}}{1+e^{(-0.951)}+e^{(-1.250)}}=0.231=P(Y=1\,|\,X=0)=\hat\pi_1(0)$. Notice also that

$\log\left[\dfrac{0.432/0.318}{0.231/0.598}\right]=1.2575=\hat\beta_{11}$. Similarly, $\hat\pi_2(1)=0.250$ and $\hat\pi_2(0)=0.171$, so that

$\log\left[\dfrac{0.250/0.318}{0.171/0.598}\right]=1.0113\approx\hat\beta_{21}$.

Finally, the almost parallel lines in the graph above imply that the odds ratios when `ME==1` and `ME==2` are approximately equal.

Testing the equality of the two odds ratios

To test whether $OR_1 = OR_2$ is the same as 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
```

The odds ratio is $\hat{\Psi} = \dfrac{(85)(11)}{(63)(19)} = 0.781$. Notice that $\hat{\Psi} = e^{(\hat{\beta}_{21} - \hat{\beta}_{11})} = e^{(1.009 - 1.256)} = e^{-0.247}$. Its standard deviation is $\hat{\sigma} = \sqrt{\frac{1}{85} + \frac{1}{11} + \frac{1}{63} + \frac{1}{19}} = 0.4137$. The test statistic is $z = \dfrac{\ln(\hat{\Psi})}{\hat{\sigma}} = -0.597$, which is associated with a p value p=0.551. There is no significant difference between the two odds ratios.

This is a consistent result to the graph above.

Testing the equality of the two odds ratios (continued)

The null hypothesis $H_o : \Psi = 1 \equiv H_o : \beta_{21} = \beta_{11} \equiv H_o : \beta_{21} - \beta_{11} = 0$. This can be tested as follows:

```
. 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
```

$H_0: \beta_{11} = 0$

$H_0: \beta_{21} = 0$

$H_0: \beta_{11} = \beta_{21}$

Notice that the chi-square statistic 0.36 is equal to the square of the *z* statistic mentioned above.

Testing of hypotheses (continued)

The significance of the effect of family history of breast cancer (`hist`) can be measured by the Wald statistic (with two degrees of freedom) or, preferably, the likelihood-ratio test. This test is derived from the comparison of the model containing the factor `hist` versus the null model.

```
. 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)

. lrtest, saving(0)
```

In the null model, the estimates $\hat{\beta}_1 = -0.811 = \log(104/234)$ while $\hat{\beta}_2 = -1.151 = \log(74/234)$.

Likelihood-ratio and Wald tests

The likelihood-ratio test is produced as follows:

```
. lrtest, using(1) model(0)
Mlogit:  likelihood-ratio test                          chi2(2)      =     12.86
                                                        Prob > chi2  =    0.0016
```

The LR test can also be derived manually as $-2\log\lambda = -2(-402.599-(-396.170)) = 12.86$.

On the other hand, the Wald test is given as follows:

```
. 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
```

$$H_0 : \begin{pmatrix} \beta_{11} \\ \beta_{21} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Both statistical tests reach the same conclusion: History of breast cancer is a significant predictive factor with respect to the frequency of Mamograms. Note that the degrees of freedom are (3-1)x1.

BE: (Επίπεδα εξερτημένης-1)*αριθμός ανεξάρτητων μεταβλητών

Incorporating a polytomous covariate

We will consider the addition of a categorical covariate with more than two categories. In the mammography example above, we investigate the significance of factor `detc` ("How likely is it that a mammogram will detect a new case of breast cancer"). The 3×3 contingency table that corresponds to this problem is as follows:

```
. 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
```

Effect of in detc predicting frequency of mammograms

```
. 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)
```

The reference category of is detc==1, that is, “Not likely”. The LR test corresponding to the overall significance of the effect of a woman’s opinion on the effectiveness of mammography on her decision to have a mammogram.

Interpretation of the estimated coefficients (logit I)

Two design variables `Idetc_1` (`detc==2`) and `Idetc_3` (`detc==3`) have been created. The estimated coefficients from the logistic regression are as follows:

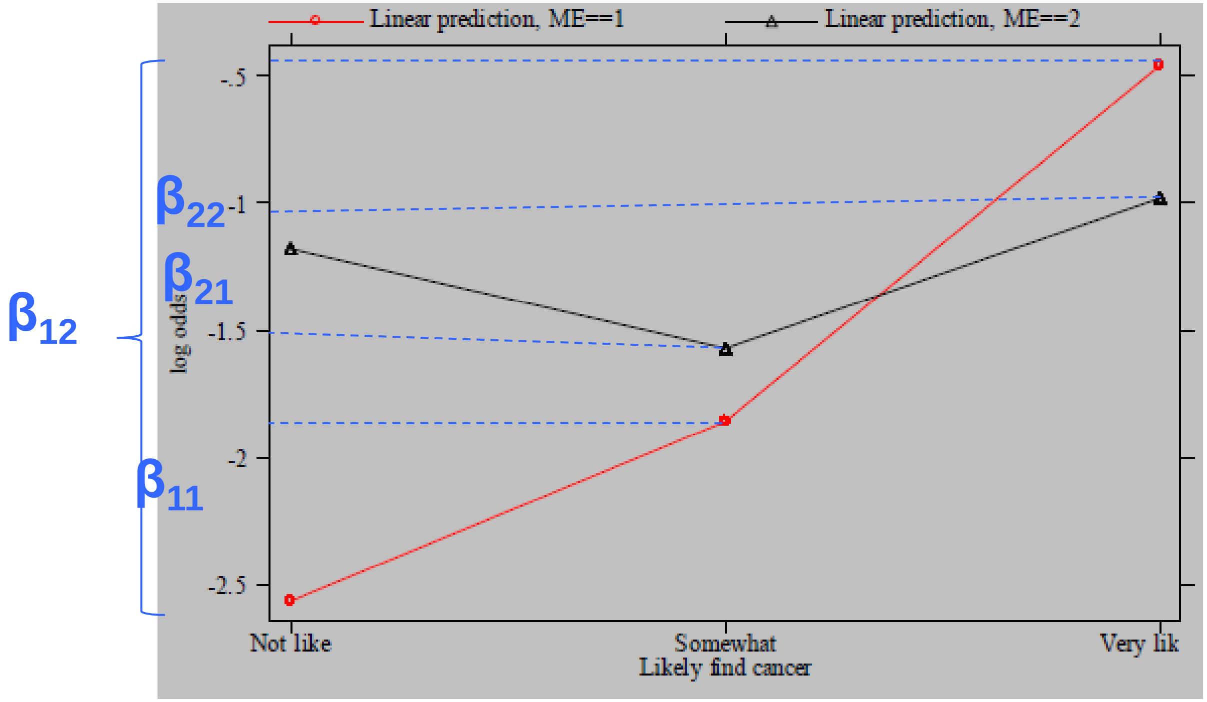
- $\hat{\beta}_{11} = 0.706 = \log\left[\frac{(13)(12)}{(77)(1)}\right] = \log(2.026)$. That is, women who think that mammograms are "somewhat likely" to detect new breast cancers are more than twice as likely (since the odds ratio estimate is $\hat{\Psi} = 2.026$) to have had a mammogram within one year, compared to women that think that mammograms are "not likely" to detect new cases. Notice that $\hat{\beta}_{11}$ is not significantly different from zero (p-value=0.652).
- Similarly, $\hat{\beta}_{12} = 2.106 = \log\left[\frac{(13)(91)}{(144)(1)}\right] = \log(8.215)$, meaning that women who think mammograms are "very likely" to discover new cases of breast cancer are more than eight times more likely ($\hat{\Psi} = 8.215$) to have had a mammogram within the past year compared to women that consider mammograms "not likely" to detect cancer. $\hat{\beta}_{12}$ is significantly different from zero (p=0.044).

Interpretation of the estimated coefficients (logit II)

- $\hat{\beta}_{21} = -0.393 = \log\left[\frac{(13)(16)}{(77)(4)}\right] = \log(0.675)$. That is, women who think that mammograms are "somewhat likely" to detect new breast cancers are 1.5 times (=1/0.675) *less* likely (since the odds ratio estimate is $\hat{\Psi} = 0.675$) to have had a mammogram over one year ago, compared to women that think that mammograms are "not likely" to detect new cases. $\hat{\beta}_{21}$ is not significantly different from zero (p-value 0.536).
- Similarly, $\hat{\beta}_{22} = 0.198 = \log\left[\frac{(13)(54)}{(144)(4)}\right] = \log(1.219)$, meaning that women who think mammograms are "very likely" to discover new cases of breast cancer are 22% more likely ($\hat{\Psi} = 1.219$) to have had a mammogram over one year ago compared to women that consider mammograms "not likely" to detect cancer. $\hat{\beta}_{12}$ is not significantly different from zero (p-value=0.739).

Graphical inspection of the model

```
. graph logit_1 logit_2 detc, c(lll) xlab(1 2 3) ylab border l1
> ("log odds") gap(4)
```



there is an indication of an interaction that there is a difference between the odds ratios.

Hypothesis tests

The hypothesis $H_o: \beta_{21} - \beta_{11} = 0$ and $\beta_{22} - \beta_{12} = 0$ can be tested as follows:

```
. test [1]

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

           chi2(  2) =   20.41
         Prob > chi2 =    0.0000
```

$$H_0: \begin{pmatrix} \beta_{11} \\ \beta_{12} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

```
. 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
```

$$H_0: \begin{pmatrix} \beta_{21} \\ \beta_{22} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

```
. 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
```

$$H_0: \begin{pmatrix} \beta_{11} - \beta_{21} \\ \beta_{12} - \beta_{22} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

This is a chi-square test with 2 degrees of freedom. The results are consistent with the graph (i.e., the association is strongest when comparing the women who have had a mammograph within the last year to those who had never had one, and comparing the not likely to very likely response).

Assessment of the significance of a continuous factor

We can measure the effect of perceived benefit (pb) of mammography (higher scores denote a smaller perceived benefit). The output of the STATA command `mlogit` is as follows:

```
. 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)
```

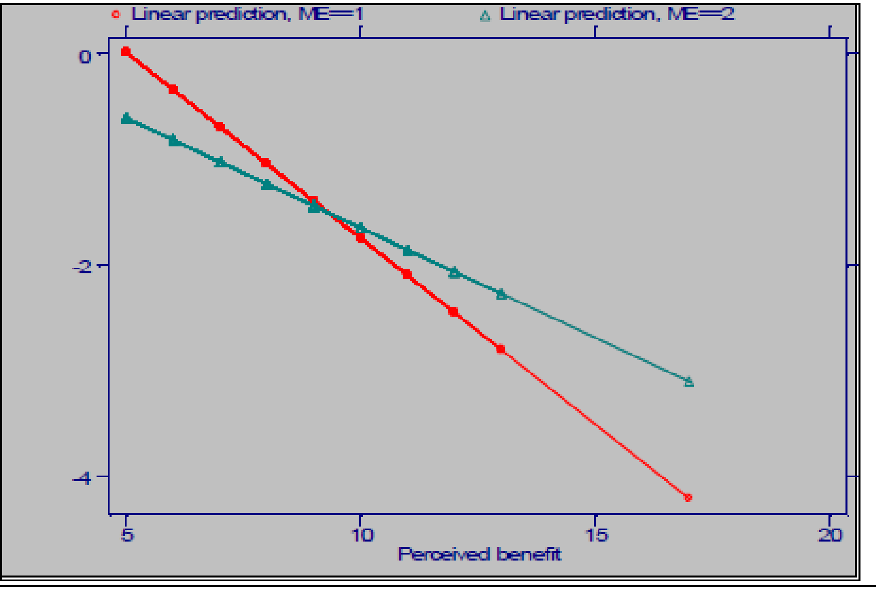
0.70 (1.42)

We can see graphically the model as follows:

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

. predict lhat2, xb outcome(2)

. graph lhat* pb, xlab ylab c(ll)  border
```



The impression from the graph is that there is a differential relationship between perceived benefit and the odds of having a mammogram. Overall, the lower the perceived benefit the lower the probability of a mammogram.

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
```

$$H_0 : \begin{pmatrix} \beta_{11} \\ \beta_{21} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

On the other hand, we can test whether the relationship between pb and ME as follows:

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

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

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

$$H_0 : \beta_{11} = \beta_{21}$$

We see that the result of this statistical test (chi-square with one degree of freedom) does not totally reflect the graphical picture in the previous page but implies a certain difference in the two groups ("within one year" and "over a year ago") in terms of the impact of the perceived benefit on the probability of mammogram.

The method of Begg & Gray (Biometrika, 1984)

Begg and Gray suggest that multinomial logistic regression can be fit by separately fitting *k*-1 logistic regressions (where *k* are the levels of the outcome variable). Note that in accordance to the analysis in Hosmer & Lemeshow (page 275-279) variable `sympt` has been dichotomized as "Agree/Strongly agree" versus "Disagree/Strongly disagree" and variable `detc` has been dichotomized as "Not likely/Somewhat likely" versus "Very likely".

We fit two logistic regressions, one for `ME==1|ME==0` and one for `ME==2|ME==0` as follows:

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

Log likelihood = -161.78145

------------------------------------------------------------------------------
      ME |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
Isympt_1 |  -2.091002   .4651287    -4.496   0.000      -3.002638   -1.179366
      pb |  -.2426146    .073756    -3.289   0.001      -.3871737   -.0980554
 Ihist_1 |   1.385025   .4682596     2.958   0.003       .4672527    2.302796
  IBSE_1 |   1.363308   .5338994     2.553   0.011       .3168847    2.409732
Idetcd_1 |   -.852694   .3654564    -2.333   0.020      -1.568975   -.1364125
   _cons |   .1786085   .7400723     0.241   0.809      -1.271907    1.629124
------------------------------------------------------------------------------
```

The method of Begg & Gray (continued)

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

Log likelihood = -153.47232                        Pseudo R2       =     0.0963

------------------------------------------------------------------------------
          ME |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
-------------+----------------------------------------------------------------
   Isympt_1 |   -1.15299   .3565788    -3.233   0.001       -1.851871   -.4541082
         pb |  -.1537696   .0726013    -2.118   0.034       -.2960655   -.0114736
    Ihist_1 |   1.097696   .4593413     2.390   0.017        .1974035    1.997988
     IBSE_1 |   .9534998   .5097419     1.871   0.061       -.0455759    1.952576
   Idetcd_1 |  -.0987046   .3190788    -0.309   0.757       -.7240876    .5266785
      _cons |  -.5864061    .744739    -0.787   0.431       -2.046068    .8732556
```

The advantage of the method of Begg and Gray is that model selection and checking can proceed individually in each of the subgroups, a greatly simplified process compared to the multinomial logistic case.

The full multinomial logistic regression is as follows (we follow Hosmer and Lemeshow's analysis from Table 8.10 on page 279):

```
Log likelihood =  -349.5663

------------------------------------------------------------------------------
      ME |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
Within o |
Isympt_1 |   -2.09475   .4574302     -4.579   0.000      -2.991297   -1.198203
      pb |  -.2494746    .072579     -3.437   0.001      -.3917269   -.1072224
 Ihist_1 |   1.309864   .4336022      3.021   0.003       .4600195    2.159709
  IBSE_1 |   1.237011   .5254241      2.354   0.019        .207199    2.266824
Idetcd_1 |  -.8851839   .3562379     -2.485   0.013      -1.583397   -.1869705
   _cons |   .3561754   .7340069      0.485   0.628      -1.082452    1.794802
---------+--------------------------------------------------------------------
Over a y |
Isympt_1 |  -1.127417   .3563621     -3.164   0.002      -1.825874   -.4289603
      pb |  -.1543182   .0726206     -2.125   0.034       -.296652   -.0119845
 Ihist_1 |   1.063179   .4528412      2.348   0.019       .1756263    1.950731
  IBSE_1 |   .9560104   .5073366      1.884   0.060      -.0383511    1.950372
Idetcd_1 |  -.1141572   .3182122     -0.359   0.720      -.7378416    .5095272
   _cons |  -.5823074   .7412705     -0.786   0.432      -2.035171    .8705562
------------------------------------------------------------------------------
(Outcome ME==Never is the comparison group)
```

The results are very close to those shown above from the individually-fit logistic regressions

Checking the goodness-of fit: The Hosmer and Lemeshow test

```
. quietly xi: logit ME i.symptd pb i.hist BSE i.detcd if ME==1|ME==0

. lfit, group(10)

Logistic model for ME, goodness-of-fit test
(Table collapsed on quantiles of estimated probabilities)

       number of observations =       338
             number of groups =        10
      Hosmer-Lemeshow chi2(8) =        12.20
                  Prob > chi2 =         0.1424

.  quietly xi: logit ME i.symptd pb i.hist BSE i.detcd if ME==2|ME==0

. lfit, group(10)

Logistic model for ME, goodness-of-fit test
(Table collapsed on quantiles of estimated probabilities)

       number of observations =       308
             number of groups =        10
      Hosmer-Lemeshow chi2(8) =         9.62
                  Prob > chi2 =         0.2929
```

The H-L statistics show that the model fits adequately.

Checking the goodness-of fit: The Pearson chi-square statistic

```
.  quietly xi: logit ME i.symptd pb i.hist BSE i.detcd if ME==1|ME==0

. lfit

Logistic model for ME, goodness-of-fit test

       number of observations =       338
 number of covariate patterns =        74
             Pearson chi2(68) =        67.84
                  Prob > chi2 =         0.4828

.  quietly xi: logit ME i.symptd pb i.hist BSE i.detcd if ME==2|ME==0

. lfit

Logistic model for ME, goodness-of-fit test

       number of observations =       308
 number of covariate patterns =        75
             Pearson chi2(69) =        63.83
                  Prob > chi2 =         0.6535
```

Note the degrees of freedom. They are equal to k-(p+1). So in the first model they are 74-(5+1)=68, while in the second model they are 75-(5+1)=69. The p values support the good fit of the model.

Checking the goodness-of fit: The Stukel test* (*JASA*, 1988)

The Stukel test is implemented as follows (Hosmer & Lemeshow, page 155):

Step 1: Produce the predicted probabilities $\hat{\pi}_j$, j=1,…,k over all covariate patterns k.

Step 2. Produce the fitted logits $\hat{g}_j = \log\left(\frac{\hat{\pi}_j}{1-\hat{\pi}_j}\right) = \mathbf{x}'_j\hat{\boldsymbol{\beta}}$, $j=1,\ldots,k$ over all covariate patterns k.

Step 3. Compute two new covariates $z_{1j} = -0.5 \times \hat{g}_j^2 \times \mathrm{I}(\hat{\pi}_j \geq 0.5)$ and $z_{2j} = (0.5) \times \hat{g}_j^2 \times \mathrm{I}(\hat{\pi}_j < 0.5)$

Step 4. Perform the Score test for the addition of z_1 and z_2 into the model. Alternatively, we can perform the likelihood-ratio test.

---

* Optional topic

Checking the goodness-of fit: The Stukel test (continued)

```
. quietly xi: logit ME i.symptd pb i.hist BSE i.detcd if ME==1|ME==0
. lrtest, saving(10)
. predict phat1
(option p assumed; Pr(ME))
. gen g1=log(phat1/(1-phat1))
. gen z11=.5*(g1)^2*(phat1>=0.5)
. gen z21=-.5*(g1)^2*(phat1<0.5)
. quietly xi: logit ME i.symptd pb i.hist i.BSE i.detcd z11 z21 if ME==1|ME==0
. lrtest, saving(11)
. lrtest, using(11) model(10)
Logit:  likelihood-ratio test                        chi2(2)      =       1.02
                                                     Prob > chi2 =     0.6006


. quietly xi: logit ME i.symptd pb i.hist BSE i.detcd if ME==2|ME==0
. lrtest, saving(20)
. predict phat2
(option p assumed; Pr(ME))
. gen g2=log(phat2/(1-phat2))
. gen z21=.5*(g2)^2*(phat2>=0.5)
. gen z22=-.5*(g2)^2*(phat2<0.5)
. quietly xi: logit ME i.symptd pb i.hist i.BSE i.detcd z21 z22 if ME==2|ME==0
. lrtest, saving(21)
. lrtest, using(21) model(20)
. lrtest, using(21) model(20)
Logit:  likelihood-ratio test                        chi2(2)      =       1.86
                                                     Prob > chi2 =     0.3937
```

The Stukel test supports the adequate fit of both models.