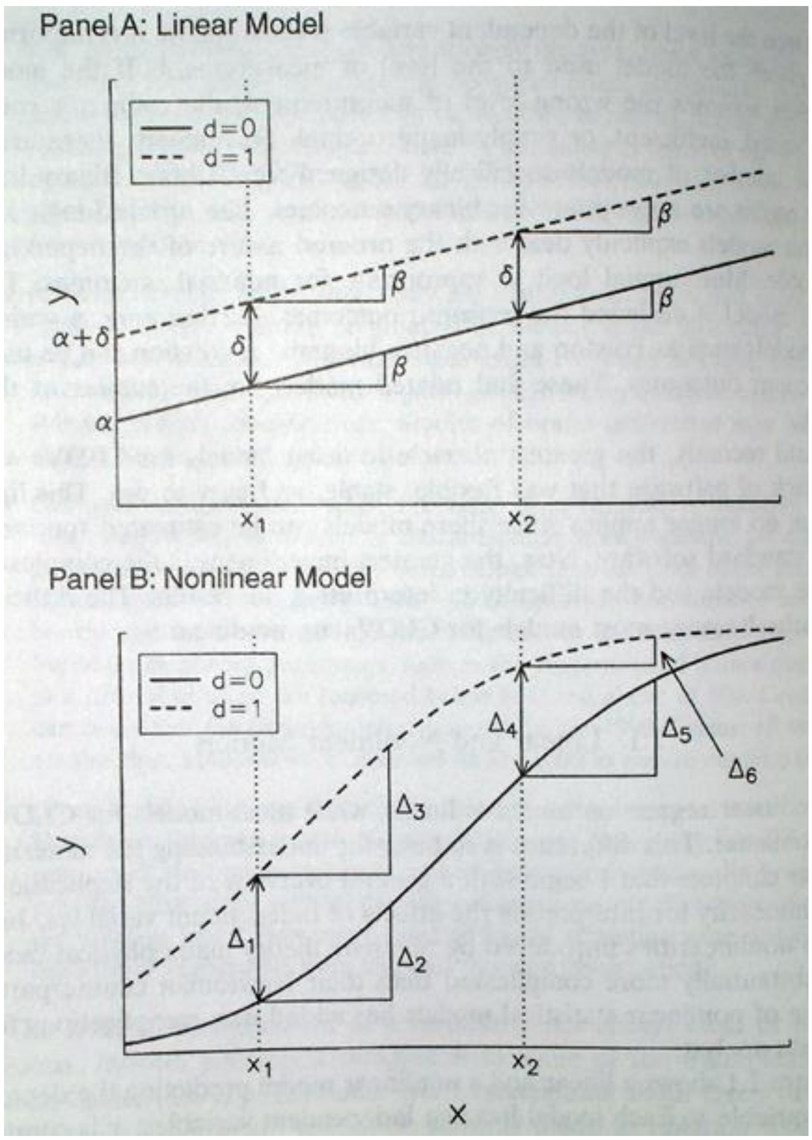
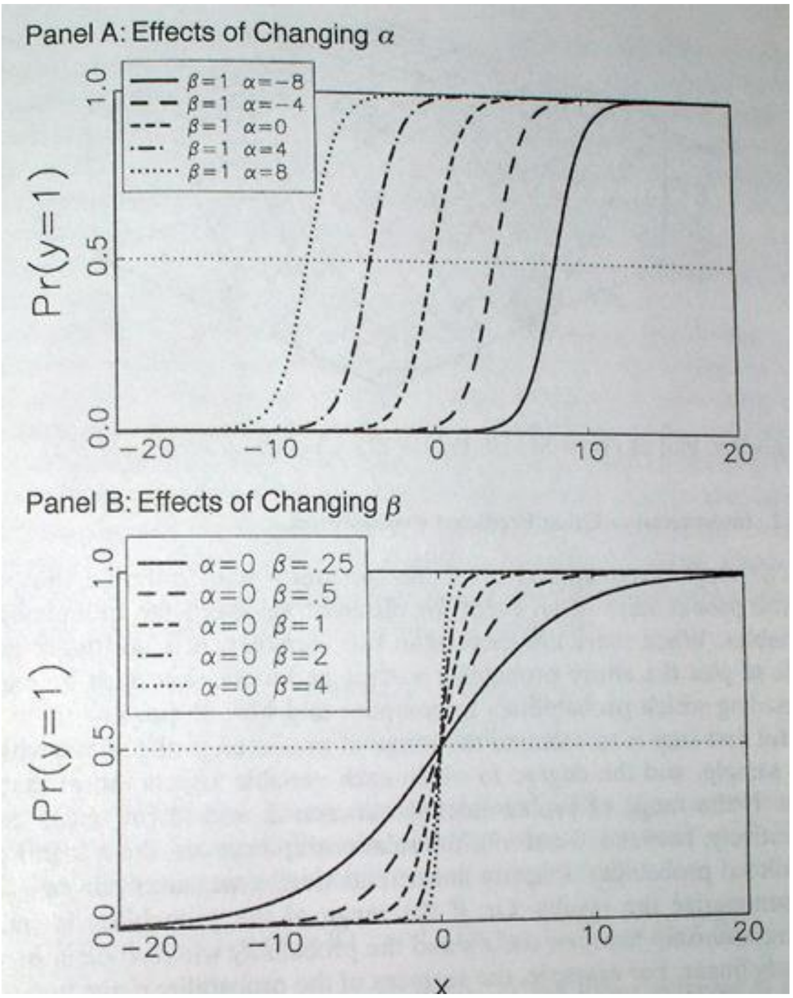
**Sociology 7704: Regression Models for Categorical Data**
**Instructor: Natasha Sarkisian**

## Binary Logit: Interpretation

As logistic regression models (whether binary, ordered, or multinomial) are nonlinear, they pose a challenge for interpretation. The increase in the dependent variable in a linear model is constant for all values of X. Not so for logit models – probability increases or decreases per unit change in X is nonconstant, as illustrated in this picture.




When interpreting logit regression coefficients, we can interpret only the sign and significance of the coefficients – cannot interpret the size. The following picture can give you an idea how the shape of the curve varies depending on the size of the coefficient, however. Note that, similarly to OLS regression, the constant determines the position of the curve along the X axis and the coefficient (beta) determines the slope.




Next, we'll examine various ways to interpret logistic regression results.

<u>1. Coefficients and Odds Ratios</u>

We'll use another model, focusing now on the probability of voting.

```
. codebook vote00
--------------------------------------------------------------------------------
vote00
did r vote in 2000 election
--------------------------------------------------------------------------------
                  type:  numeric (byte)
                 label:  vote00

                 range:  [1,4]                        units:  1
         unique values:  4                        missing .:  14/2765

            tabulation:  Freq.   Numeric  Label
                          1780         1  voted
                           822         2  did not vote
                           138         3  ineligible
                            11         4  refused to answer
                            14         .

. gen vote=(vote00==1) if vote00<3
(163 missing values generated)
. gen married=(marital==1)
```

```
. logit vote age sex born married childs educ
Iteration 0:   log likelihood = -1616.8899
Iteration 1:   log likelihood = -1365.9814
Iteration 2:   log likelihood = -1353.4091
Iteration 3:   log likelihood = -1353.2224
Iteration 4:   log likelihood = -1353.2224
Logistic regression                               Number of obs   =       2590
                                                  LR chi2(6)      =     527.33
                                                  Prob > chi2     =     0.0000
Log likelihood = -1353.2224                       Pseudo R2       =     0.1631
------------------------------------------------------------------------------
        vote |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0466321    .003337    13.97   0.000     .0400917    .0531726
         sex |   .1094233     .09552     1.15   0.252    -.0777924     .296639
        born |  -.9673683   .1859278    -5.20   0.000     -1.33178   -.6029564
     married |   .4911099   .0983711     4.99   0.000     .2983062    .6839136
      childs |  -.0391447   .0327343    -1.20   0.232    -.1033028    .0250133
        educ |   .2862839   .0197681    14.48   0.000     .2475391    .3250287
       _cons |  -4.352327   .3892601   -11.18   0.000    -5.115263   -3.589391
------------------------------------------------------------------------------
```

These are regular logit coefficients; so we can interpret the sign and significance but not the size of effects. So we can say that age increases the probability of voting but we can't say by how much – that's because a 1 year increase in age will not affect the probability the same way for a 30 year old and for a 40 year old.

To be able to interpret effect size, we turn to odds ratios. Note that odds ratios are only appropriate for logistic regression – they don't work for probit models.

Odds are ratios of two probabilities – probability of a positive outcome and a probability of a negative outcome (e.g. probability of voting divided by a probability of not voting). But since probabilities vary depending on values of X, such a ratio varies as well. What remains constant is the ratio of such odds – e.g. odds of voting for women divided by odds of voting for men will be the same number regardless of the values of other variables. Similarly, the odds ratio for age can be a ratio of the odds of voting for someone who is 31 y.o. to the odds of a 30 y.o. person, or of a 41 y.o. to a 40 y.o. person's odds – these will be the same regardless of what age values you pick, as long as they are one year apart. So let's examine the odds ratios.

```
. logit vote age sex born married childs educ, or
Iteration 0:   log likelihood = -1616.8899
Iteration 1:   log likelihood = -1365.9814
Iteration 2:   log likelihood = -1353.4091
Iteration 3:   log likelihood = -1353.2224
Iteration 4:   log likelihood = -1353.2224
Logistic regression                               Number of obs   =       2590
                                                  LR chi2(6)      =     527.33
                                                  Prob > chi2     =     0.0000
Log likelihood = -1353.2224                       Pseudo R2       =     0.1631
------------------------------------------------------------------------------
        vote | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   1.047736   .0034963    13.97   0.000     1.040906    1.054612
         sex |   1.115634   .1065654     1.15   0.252     .9251564     1.34533
        born |    .380082   .0706678    -5.20   0.000     .2640069    .5471915
     married |   1.634129    .160751     4.99   0.000     1.347574    1.981618
      childs |   .9616115   .0314777    -1.20   0.232     .9018538    1.025329
        educ |    1.33147   .0263207    14.48   0.000     1.280869     1.38407
------------------------------------------------------------------------------
```

Another way to obtain odds ratios would be to use "logistic" command instead of "logit" – it automatically displays odds ratios instead of coefficients. But yet another, more convenient way is to use listcoef command (that's one of the commands written by Scott Long that we downloaded as a part of spost package):

```
. listcoef
logit (N=2590): Factor change in odds
  Odds of: 1 vs 0
----------------------------------------------------------------------
             |        b         z     P>|z|       e^b    e^bStdX     SDofX
-------------+--------------------------------------------------------
         age |    0.0466   13.974    0.000     1.048      2.230    17.195
         sex |    0.1094    1.146    0.252     1.116      1.056     0.497
        born |   -0.9674   -5.203    0.000     0.380      0.788     0.246
     married |    0.4911    4.992    0.000     1.634      1.278     0.499
      childs |   -0.0391   -1.196    0.232     0.962      0.936     1.676
        educ |    0.2863   14.482    0.000     1.331      2.311     2.926
    constant |   -4.3523  -11.181    0.000         .          .         .
----------------------------------------------------------------------
```

The advantage of listcoef is that it reports regular coefficients, odds ratios, and standardized odds ratios in one table. Odds ratios are exponentiated logistic regression coefficients. They are sometimes called factor coefficients, because they are multiplicative coefficients. Odds ratios are equal to 1 if there is no effect, smaller than 1 if the effect is negative and larger than 1 if it is positive. So for example, the odds ratio for married indicates that the odds of voting for those who are married are 1.63 times higher than for those who are not married. And the odds ratio for education indicates that each additional year of education makes one's odds of voting 1.33 times higher -- or, in other words, increases those odds by 33%. To get percent change directly, we can use percent option:

```
. listcoef, percent
logit (N=2590): Percentage Change in Odds
  Odds of: 1 vs 0
-------------------------------------------------------------------
        vote |      b         z      P>|z|      %      %StdX      SDofX
-------------+-----------------------------------------------------
         age |   0.04663   13.974   0.000      4.8    123.0    17.1953
         sex |   0.10942    1.146   0.252     11.6      5.6     0.4972
        born |  -0.96737   -5.203   0.000    -62.0    -21.2     0.2457
     married |   0.49111    4.992   0.000     63.4     27.8     0.4990
      childs |  -0.03914   -1.196   0.232     -3.8     -6.4     1.6762
        educ |   0.28628   14.482   0.000     33.1    131.1     2.9257
-------------------------------------------------------------------
```

Beware: if you would like to know what the increase would be per, say, 10 units increase in the independent variable – e.g. 10 years of education, you cannot simply multiple the odds ratio by 10! The coefficient, in fact, would be odds ratio to the power of 10. Or alternatively, you could take the regular logit coefficient, multiply it by 10 and then exponentiate it -- e.g., for education:

```
. di exp(0.28628*10)
17.510488
. di 1.3315^10
17.515063
```

Standardized odds ratios (presented under e^bStdX) are similar to regular odds ratios, but they display the change in the odds of voting per one standard deviation change in the independent variable. The last column in the table generated by listcoef shows what one standard deviation for each variable is. So for age the standardized odds ratio indicates that 17 years of age increase one's odds of voting 2.23 times, or by 123%. Standardized odds ratios, like standardized coefficients in OLS, allow us to compare effect sizes across variables regardless of their measurement units. But, beware of comparing negative and positive effects – odds ratios of 1.5 and .5 are not equivalent, even though the first one represents a 50% increase in odds and the second one represents a 50% decrease. This is because odds ratios cannot be below zero (there cannot be a decrease more than 100%), but they do not have an upper bound – i.e. can be infinitely high. In order to be able to compare positive and negative effects, we can reverse odds ratios and generate odds ratios for odds of not voting (rather than odds of voting).

```
. listcoef, reverse
logit (N=2590): Factor Change in Odds
  Odds of: 0 vs 1
----------------------------------------------------------------------
        vote |      b         z     P>|z|    e^b    e^bStdX      SDofX
-------------+--------------------------------------------------------
         age |   0.04663   13.974   0.000   0.9544   0.4485    17.1953
         sex |   0.10942    1.146   0.252   0.8964   0.9470     0.4972
        born |  -0.96737   -5.203   0.000   2.6310   1.2682     0.2457
     married |   0.49111    4.992   0.000   0.6119   0.7826     0.4990
      childs |  -0.03914   -1.196   0.232   1.0399   1.0678     1.6762
        educ |   0.28628   14.482   0.000   0.7510   0.4328     2.9257
```

We can see for example that the odds ratio of 0.3801 for born is a negative effect corresponding in size to a positive odds ratio of 2.6310. Listcoef also has a help option that explains what's what :

```
. listcoef, reverse help
logit (N=2590): Factor Change in Odds
  Odds of: 0 vs 1
----------------------------------------------------------------------
        vote |      b         z     P>|z|    e^b    e^bStdX      SDofX
-------------+--------------------------------------------------------
         age |   0.04663   13.974   0.000   0.9544   0.4485    17.1953
         sex |   0.10942    1.146   0.252   0.8964   0.9470     0.4972
        born |  -0.96737   -5.203   0.000   2.6310   1.2682     0.2457
     married |   0.49111    4.992   0.000   0.6119   0.7826     0.4990
      childs |  -0.03914   -1.196   0.232   1.0399   1.0678     1.6762
        educ |   0.28628   14.482   0.000   0.7510   0.4328     2.9257
----------------------------------------------------------------------
       b = raw coefficient
       z = z-score for test of b=0
   P>|z| = p-value for z-test
     e^b = exp(b) = factor change in odds for unit increase in X
 e^bStdX = exp(b*SD of X) = change in odds for SD increase in X
   SDofX = standard deviation of X
```

When a set of dummies is used, we might be interested in all kinds of pairwise comparisons; to get odds ratios for those, we use pwcompare command:

```
. logit vote age sex born i.marital childs educ, or
Iteration 0:   log likelihood = -1616.8899
Iteration 1:   log likelihood = -1361.6039
Iteration 2:   log likelihood = -1352.4837
Iteration 3:   log likelihood = -1352.4548
Iteration 4:   log likelihood = -1352.4548
Logistic regression                               Number of obs   =       2590
                                                  LR chi2(9)      =     528.87
                                                  Prob > chi2     =     0.0000
Log likelihood = -1352.4548                       Pseudo R2       =     0.1635
------------------------------------------------------------------------------
        vote | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   1.048782   .0040525    12.33   0.000     1.040869    1.056755
         sex |    1.11771   .1080131     1.15   0.250      .924849    1.350789
        born |   .3761262   .0701482    -5.24   0.000     .2609655    .5421061
             |
     marital |
```

```
      widowed  |   .6014296    .125745    -2.43   0.015     .3992255    .9060482
     divorced  |   .5493787   .0741513    -4.44   0.000     .4216796    .7157496
    separated  |   .6970315   .1716079    -1.47   0.143     .4302175    1.129319
never married  |   .6503118   .0840993    -3.33   0.001     .5047112    .8379156
               |
        childs |   .9655952   .0325389    -1.04   0.299     .9038806    1.031523
          educ |   1.333732   .0265289    14.48   0.000     1.282737    1.386754
         _cons |   .0196952   .0081736    -9.46   0.000     .0087319    .0444234
--------------------------------------------------------------------------------
```

```
. pwcompare marital
Pairwise comparisons of marginal linear predictions
Margins      : asbalanced
-----------------------------------------------------------------------------
                            |                                 Unadjusted
                            |   Contrast   Std. Err.     [95% Conf. Interval]
----------------------------+------------------------------------------------
vote                        |
                    marital |
        widowed vs married  |  -.5084458   .2090768     -.9182288   -.0986628
       divorced vs married  |  -.5989672   .1349731     -.8635096   -.3344249
      separated vs married  |  -.3609247   .2461983     -.8434645     .121615
  never married vs married  |  -.4303034   .1293215     -.6837689   -.1768379
       divorced vs widowed  |  -.0905214   .2213725     -.5244036    .3433607
      separated vs widowed  |   .1475211   .3044299     -.4491506    .7441927
  never married vs widowed  |   .0781424   .2412223     -.3946447    .5509295
     separated vs divorced  |   .2380425   .2618905     -.2752534    .7513384
 never married vs divorced  |   .1686638   .1560947     -.1372761    .4746038
never married vs separated  |  -.0693787   .2594929     -.5779754    .4392181
-----------------------------------------------------------------------------
```

And to get actual odds ratios:

```
. pwcompare marital, eform
Pairwise comparisons of marginal linear predictions
Margins      : asbalanced
-----------------------------------------------------------------------------
                            |                                 Unadjusted
                            |     exp(b)   Std. Err.     [95% Conf. Interval]
----------------------------+------------------------------------------------
vote                        |
                    marital |
        widowed vs married  |   .6014296    .125745      .3992255    .9060482
       divorced vs married  |   .5493787   .0741513      .4216796    .7157496
      separated vs married  |   .6970315   .1716079      .4302175    1.129319
  never married vs married  |   .6503118   .0840993      .5047112    .8379156
       divorced vs widowed  |   .9134548   .2022138      .5919083    1.409677
      separated vs widowed  |   1.158958   .3528214        .63817    2.104742
  never married vs widowed  |   1.081277   .2608281      .6739194    1.734865
     separated vs divorced  |   1.268763    .332277      .7593797    2.119835
 never married vs divorced  |   1.183722   .1847727      .8717295    1.607377
never married vs separated  |   .9329733      .2421      .5610331    1.551494
-----------------------------------------------------------------------------
```

A side note: something that can be helpful when doing hypothesis testing for groups of dummies (instead of using acc option in test or lrtest):

```
. testparm i.marital
 ( 1)  [vote]2.marital = 0
 ( 2)  [vote]3.marital = 0
 ( 3)  [vote]4.marital = 0
 ( 4)  [vote]5.marital = 0

           chi2(  4) =   26.50
         Prob > chi2 =    0.0000
```

2. Predicted Probabilities

In addition to regular coefficients and odds ratios, we also should examine predicted probabilities – both for the actual observations in our data and for strategically selected hypothetical cases. Predicted probabilities are always calculated for a specific set of independent variables' values. One thing we can calculate is predicted probabilities for the actual data that we have – for each case, we take the values of all independent variables and plug it into the equation:

```
. predict prob
(option p assumed; Pr(vote))
(26 missing values generated)

. sum prob if e(sample)
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        prob |      2590    .6833977     .204702   .0205784   .9926677
```

Mean of predicted probabilities represents the average proportion in the sample:

```
. sum vote if e(sample)
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        vote |      2590    .6833977    .4652406          0          1
```

These are predicted probabilities for the actual cases in our dataset. It can be useful, however, to calculate predicted probabilities for hypothetical sets of values – some interesting combinations that we could compare and contrast.

```
. margins, atmeans

Adjusted predictions                              Number of obs   =       2590
Model VCE    : OIM

Expression   : Pr(vote), predict()
at           : age             =    46.93591 (mean)
               sex             =    1.553282 (mean)
               born            =    1.064479 (mean)
               married         =    .4675676 (mean)
               childs          =    1.838996 (mean)
               educ            =    13.39459 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .7249026   .0100274    72.29   0.000     .7052494    .7445559
------------------------------------------------------------------------------
```

This calculates a predicted probability for a case with all values set at the mean. So an "average" person has 72.5% chance of voting. We can also see what these averages are. If we do not specify atmeans (and do not specify values for each variable), the margins command calculates average predicted probability across the observations we have in the dataset.

Clearly, for some variables, averages don't make sense – e.g., we don't want to use averages for dummy variables; rather, we'd want to specify what values to use. Here is an example of specifying values:

```
. margins, at(age=30 born=1 sex=2 married=0) atmeans

Adjusted predictions                              Number of obs   =       2590
Model VCE    : OIM
```

```
Expression    : Pr(vote), predict()
at            : age             =          30
                sex             =           2
                born            =           1
                married         =           0
                childs          =    1.838996 (mean)
                educ            =    13.39459 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .5151914   .0219222    23.50   0.000     .4722246    .5581581
------------------------------------------------------------------------------
```

This is the predicted value for someone who is 30, native born, female, and unmarried (and has average number of children and average education). Note that if you have a set of dummy variables, you can just specify the category number, e.g., if you are using i.marital, you can write (marital=2) in the at option.

We can also use margins command to compare predictions at different values:

```
. margins, at(married=0 married=1) atmeans
Adjusted predictions                              Number of obs   =       2590
Model VCE    : OIM
Expression   : Pr(vote), predict()
1._at         : age             =    46.93591 (mean)
                sex             =    1.553282 (mean)
                born            =    1.064479 (mean)
                married         =           0
                childs          =    1.838996 (mean)
                educ            =    13.39459 (mean)

2._at         : age             =    46.93591 (mean)
                sex             =    1.553282 (mean)
                born            =    1.064479 (mean)
                married         =           1
                childs          =    1.838996 (mean)
                educ            =    13.39459 (mean)
------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .6768395   .0143948    47.02   0.000     .6486262    .7050528
          2  |   .7738877   .0131271    58.95   0.000      .748159    .7996164
------------------------------------------------------------------------------

. margins, at(age=(30(10)70)) atmeans

Adjusted predictions                              Number of obs   =       2590
Model VCE    : OIM

Expression   : Pr(vote), predict()

1._at         : age             =          30
                sex             =    1.553282 (mean)
                born            =    1.064479 (mean)
                married         =    .4675676 (mean)
                childs          =    1.838996 (mean)
                educ            =    13.39459 (mean)
```

```
2._at        : age             =          40
               sex             =    1.553282 (mean)
               born            =    1.064479 (mean)
               married         =    .4675676 (mean)
               childs          =    1.838996 (mean)
               educ            =    13.39459 (mean)

3._at        : age             =          50
               sex             =    1.553282 (mean)
               born            =    1.064479 (mean)
               married         =    .4675676 (mean)
               childs          =    1.838996 (mean)
               educ            =    13.39459 (mean)

4._at        : age             =          60
               sex             =    1.553282 (mean)
               born            =    1.064479 (mean)
               married         =    .4675676 (mean)
               childs          =    1.838996 (mean)
               educ            =    13.39459 (mean)

5._at        : age             =          70
               sex             =    1.553282 (mean)
               born            =    1.064479 (mean)
               married         =    .4675676 (mean)
               childs          =    1.838996 (mean)
               educ            =    13.39459 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .5446694   .0160415    33.95   0.000     .5132286    .5761101
          2  |   .6559903   .0111333    58.92   0.000     .6341694    .6778113
          3  |    .752464     .01005    74.87   0.000     .7327664    .7721617
          4  |   .8289379   .0106262    78.01   0.000     .8081108    .8497649
          5  |   .8853845   .0104219    84.95   0.000      .864958    .9058111
------------------------------------------------------------------------------
```

To have a more compact legend:

```
. margins, at(age=(30(10)70) married=(0 1)) atmeans noatlegend

Adjusted predictions                              Number of obs   =       2590
Model VCE    : OIM

Expression   : Pr(vote), predict()

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .4873847   .0196449    24.81   0.000     .4488815     .525888
          2  |   .6084111   .0200184    30.39   0.000     .5691757    .6476464
          3  |   .6024896   .0157359    38.29   0.000     .5716478    .6333313
          4  |   .7123775   .0151151    47.13   0.000     .6827525    .7420025
          5  |   .7072717   .0141615    49.94   0.000     .6795157    .7350278
          6  |   .7979096   .0125434    63.61   0.000      .773325    .8224942
          7  |   .7938829   .0139495    56.91   0.000     .7665424    .8212234
          8  |   .8629015   .0111527    77.37   0.000     .8410427    .8847604
          9  |   .8599425   .0132394    64.95   0.000     .8339938    .8858911
```

```
        10  |   .9093663   .0097348    93.41   0.000     .8902865    .9284462
------------------------------------------------------------------------------

. mlistat

at() values held constant

     sex      born    childs      educ
--------------------------------------
    1.55      1.06      1.84      13.4

at() values vary

   _at |       age   married
-------+--------------------
     1 |        30         0
     2 |        30         1
     3 |        40         0
     4 |        40         1
     5 |        50         0
     6 |        50         1
     7 |        60         0
     8 |        60         1
     9 |        70         0
    10 |        70         1
```

We could also separate groups and do predictions separately (note that group-based means are used for each group, so it is different from using that variable within "at" option).

```
. margins, over(married) at(age=(30(10)70) ) atmeans noatlegend

Adjusted predictions                              Number of obs   =       2590
Model VCE    : OIM

Expression   : Pr(vote), predict()
over         : married

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 _at#married |
        1 0  |   .4787915   .0187124    25.59   0.000     .4421158    .5154673
        1 1  |   .6177066   .0203227    30.39   0.000     .5778749    .6575383
        2 0  |   .5942195   .0151977    39.10   0.000     .5644325    .6240064
        2 1  |   .7203395   .0149981    48.03   0.000     .6909437    .7497353
        3 0  |   .7000965   .0141623    49.43   0.000      .672339     .727854
        3 1  |   .8041548   .0121038    66.44   0.000     .7804318    .8278778
        4 0  |   .7881948   .0143163    55.06   0.000     .7601354    .8162543
        4 1  |   .8674719   .0105976    81.86   0.000      .846701    .8882428
        5 0  |   .8557462   .0137381    62.29   0.000       .82882    .8826724
        5 1  |   .9125447   .0092091    99.09   0.000     .8944952    .9305942
------------------------------------------------------------------------------

. mlistat

at() values vary

   _at |       age       sex      born   married    childs      educ
-------+------------------------------------------------------------
     1 |        30      1.59      1.05         0      1.53      13.2
     2 |        30      1.51      1.08         1       2.2      13.7
     3 |        40      1.59      1.05         0      1.53      13.2
```

```
    4 |       40      1.51      1.08         1       2.2      13.7
    5 |       50      1.59      1.05         0      1.53      13.2
    6 |       50      1.51      1.08         1       2.2      13.7
    7 |       60      1.59      1.05         0      1.53      13.2
    8 |       60      1.51      1.08         1       2.2      13.7
    9 |       70      1.59      1.05         0      1.53      13.2
   10 |       70      1.51      1.08         1       2.2      13.7
```

Margins command also permits us to transform our predictions and get p-values and CI for transformed version:

```
. margins, at(married=(0 1)) atmeans noatlegend expression(1-predict(pr))

Adjusted predictions                              Number of obs   =       2590
Model VCE    : OIM

Expression   : 1-predict(pr)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .3231605   .0143948    22.45   0.000     .2949472    .3513738
          2  |   .2261123   .0131271    17.22   0.000     .2003836     .251841
------------------------------------------------------------------------------
```

Or to test if predicted probability is different from, say, 0.5:

```
. margins, at(married=(0 1)) atmeans noatlegend expression(predict(pr)-.5)

Adjusted predictions                              Number of obs   =       2590
Model VCE    : OIM

Expression   : predict(pr)-.5

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .1768395   .0143948    12.28   0.000     .1486262    .2050528
          2  |   .2738877   .0131271    20.86   0.000      .248159    .2996164
------------------------------------------------------------------------------
```

We can also use mtable to obtain values of predicted probabilities for various combinations of categorical variables – but note that we need to specify what values to use for all other variables – e.g., in this case, all other variables are set at the mean.

```
. qui logit vote age sex born married childs educ
. mtable, at(born=(0 1) married=(0 1)) atmeans

Expression: Pr(vote), predict()

           |     born   married     Pr(y)
 ----------+-----------------------------
         1 |        0         0     0.854
         2 |        0         1     0.906
         3 |        1         0     0.690
         4 |        1         1     0.785
```

```
Specified values of covariates

          |      age       sex    childs      educ
----------+----------------------------------------
  Current |     46.9      1.55      1.84      13.4
```

This allows us to see that the effect of one variable depends on the level of the other – for native born individuals, marriage increases chances of voting by 9.5%, but for the foreign born, marriage increases these chances by 12.2%. We can also get confidence intervals for predictions, as well as some other statistics:

```
. mtable, at(born=(0 1) married=(0 1)) atmeans statistics(ci)

Expression: Pr(vote), predict()

          |     born   married     Pr(y)        ll        ul
----------+--------------------------------------------------
        1 |        0         0     0.854     0.804     0.905
        2 |        0         1     0.906     0.869     0.942
        3 |        1         0     0.690     0.662     0.718
        4 |        1         1     0.785     0.759     0.810

Specified values of covariates

          |      age       sex    childs      educ
----------+----------------------------------------
  Current |     46.9      1.55      1.84      13.4

. mtable, at(born=(0 1) married=(0 1)) atmeans statistics(all)

Expression: Pr(vote), predict()

          |     born   married     Pr(y)        se         z         p
----------+------------------------------------------------------------
        1 |        0         0     0.854     0.026    33.196     0.000
        2 |        0         1     0.906     0.019    48.426     0.000
        3 |        1         0     0.690     0.014    48.515     0.000
        4 |        1         1     0.785     0.013    60.182     0.000

          |       ll        ul
----------+-------------------
        1 |    0.804     0.905
        2 |    0.869     0.942
        3 |    0.662     0.718
        4 |    0.759     0.810

Specified values of covariates

          |      age       sex    childs      educ
----------+----------------------------------------
  Current |     46.9      1.55      1.84      13.4
```

You may also find an older command, prtab, useful (but note that it is not compatible with the new way to specifying dummies using i. – only works with xi: prefix in that case):

```
. prtab born married, rest(mean)
logit: Predicted probabilities of positive outcome for vote
--------------------------
was r     |
born in   |
this      |    married
country   |      0       1
```

```
----------+---------------
      yes | 0.6903  0.7846
       no | 0.4587  0.5806
--------------------------
          age        sex       born    married     childs       educ
x=  46.935907  1.5532819  1.0644788  .46756757  1.8389961  13.394595
```

With mtable, the best way to do predictions by group is to use over option:

```
. mtable, at(born=(0 1) married=(0 1)) atmeans over(sex)

Expression: Pr(vote), predict()

           |      age       sex      born   married    childs      educ
 ----------+-----------------------------------------------------------
         1 |     46.2         1         0         0      1.68      13.4
         2 |     47.5         2         0         0      1.96      13.4
         3 |     46.2         1         0         1      1.68      13.4
         4 |     47.5         2         0         1      1.96      13.4
         5 |     46.2         1         1         0      1.68      13.4
         6 |     47.5         2         1         0      1.96      13.4
         7 |     46.2         1         1         1      1.68      13.4
         8 |     47.5         2         1         1      1.96      13.4

           |    Pr(y)
 ----------+---------
         1 |    0.843
         2 |    0.863
         3 |    0.898
         4 |    0.911
         5 |    0.672
         6 |    0.705
         7 |    0.770
         8 |    0.796

Specified values where .n indicates no values specified with at()

           |  No at()
 ----------+---------
   Current |       .n
```

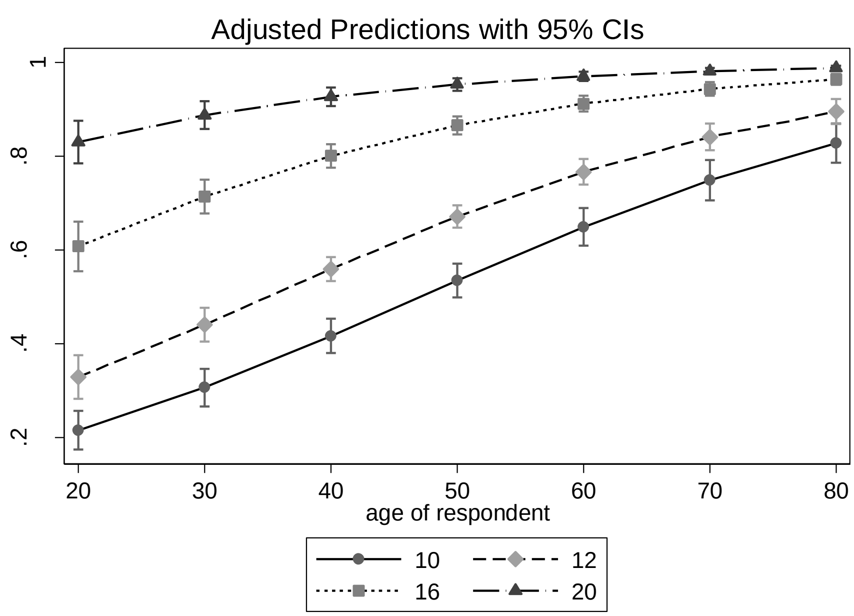
Note that it only makes sense to create such tables of predicted probabilities for variables that have significant effects – otherwise, you'll see no differences.

Further, we can use marginsplot after margin to graph probabilities for certain sets of values. This is useful with continuous variables, as it allows us to see how predicted probability changes across values of one variable (given that the rest of them are set at some specific values).

For example, we can plot four curves that show how probability of voting changes by age for an average person who has 10, 12, 16, or 20 years of education.

```
. margins, at(age=(20(10)80) educ=(10 12 16 20)) atmeans noatlegend

Adjusted predictions                              Number of obs   =       2590
Model VCE    : OIM

Expression   : Pr(vote), predict()

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
```

```
-------------+----------------------------------------------------------------
         _at |
          1  |   .2160915   .0211483    10.22   0.000     .1746416    .2575414
          2  |   .3290183   .0237367    13.86   0.000     .2824951    .3755414
          3  |   .6080911   .0271769    22.38   0.000     .5548254    .6613568
          4  |   .8307875   .0231461    35.89   0.000      .785422    .8761531
          5  |   .3074013   .0204778    15.01   0.000     .2672656    .3475371
          6  |   .4411898   .0186146    23.70   0.000     .4047058    .4776738
          7  |   .7141425   .0183676    38.88   0.000     .6781426    .7501424
          8  |   .8877053   .0151673    58.53   0.000     .8579778    .9174327
          9  |   .4167808   .0186739    22.32   0.000     .3801807    .4533809
         10  |   .5597036   .0131027    42.72   0.000     .5340229    .5853843
         11  |   .8008927   .0125282    63.93   0.000     .7763378    .8254475
         12  |   .9271563    .010058    92.18   0.000     .9074431    .9468696
         13  |   .5350154   .0185145    28.90   0.000     .4987277    .5713031
         14  |   .6717814   .0119539    56.20   0.000     .6483522    .6952105
         15  |   .8662472   .0098556    87.89   0.000     .8469306    .8855637
         16  |    .953474   .0068999   138.19   0.000     .9399504    .9669976
         17  |   .6494415    .020568    31.58   0.000     .6091288    .6897541
         18  |   .7671963   .0138781    55.28   0.000     .7399956    .7943969
         19  |   .9124937   .0084824   107.58   0.000     .8958686    .9291189
         20  |    .970585    .004878   198.97   0.000     .9610244    .9801456
         21  |   .7489234   .0220661    33.94   0.000     .7056747    .7921721
         22  |   .8414212   .0146909    57.27   0.000     .8126275    .8702149
         23  |   .9437876   .0071808   131.43   0.000     .9297136    .9578617
         24  |    .981525   .0034965   280.72   0.000      .974672     .988378
         25  |   .8276656   .0213316    38.80   0.000     .7858566    .8694747
         26  |   .8952132   .0136561    65.55   0.000     .8684477    .9219788
         27  |   .9643278   .0057912   166.52   0.000     .9529772    .9756783
         28  |   .9884446   .0025063   394.38   0.000     .9835323     .993357
------------------------------------------------------------------------------

. marginsplot

  Variables that uniquely identify margins: age educ
```

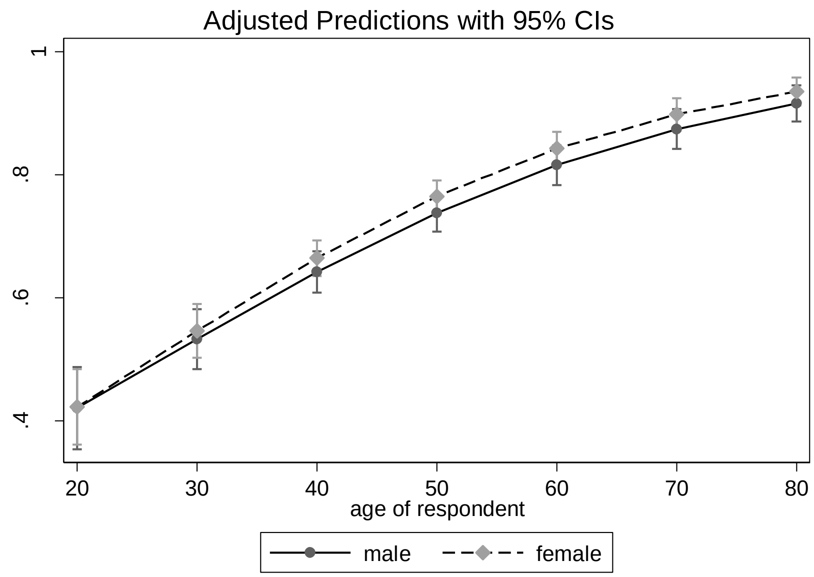



If there are interactions or nonlinearities that required that you entered a variable more than once (e.g. X and X squared), you can also this marginplots to graph that.

```
. logit vote i.sex##c.age educ i.born i.marital childs

Iteration 0:   log likelihood = -1616.8899
Iteration 1:   log likelihood = -1361.3117
Iteration 2:   log likelihood = -1352.2041
Iteration 3:   log likelihood = -1352.1752
Iteration 4:   log likelihood = -1352.1752

Logistic regression                               Number of obs   =       2590
                                                  LR chi2(10)     =     529.43
                                                  Prob > chi2     =     0.0000
Log likelihood = -1352.1752                       Pseudo R2       =     0.1637

------------------------------------------------------------------------------
        vote |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |
     female  |  -.0844048   .2788219    -0.30   0.762    -.6308857    .4620761
         age |   .0451964   .0050282     8.99   0.000     .0353414    .0550514
             |
   sex#c.age |
     female  |   .0045923    .006136     0.75   0.454     -.007434    .0166185
             |
        educ |   .2877763   .0198892    14.47   0.000     .2487942    .3267584
             |
        born |
         no  |  -.9707724   .1867578    -5.20   0.000    -1.336811   -.6047339
             |
     marital |
    widowed  |  -.5480377   .2157987    -2.54   0.011    -.9709953   -.1250801
   divorced  |  -.6021702     .13507    -4.46   0.000    -.8669025   -.3374379
  separated  |  -.3569101   .2463735    -1.45   0.147    -.8397932     .125973
never mar..  |  -.4341406   .1294304    -3.35   0.001    -.6878196   -.1804616
             |
      childs |  -.0334493   .0337876    -0.99   0.322    -.0996717    .0327732
       _cons |   -4.68753   .3754022   -12.49   0.000    -5.423305   -3.951756
------------------------------------------------------------------------------

. margins, at(age=(20(10)80) sex=(1 2 )) atmeans noatlegend

Adjusted predictions                              Number of obs   =       2590
Model VCE    : OIM

Expression   : Pr(vote), predict()

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .4208618   .0339115    12.41   0.000     .3543965    .4873271
          2  |   .5331333   .0248281    21.47   0.000     .4844712    .5817954
          3  |   .6421463   .0171949    37.35   0.000     .6084449    .6758478
          4  |   .7382043   .0153451    48.11   0.000     .7081285    .7682802
          5  |   .8158712    .016502    49.44   0.000     .7835279    .8482145
          6  |   .8744164   .0166122    52.64   0.000     .8418571    .9069757
          7  |   .9162574   .0151062    60.65   0.000     .8866497     .945865
          8  |   .4226765   .0312803    13.51   0.000     .3613681    .4839848
          9  |   .5463891    .022257    24.55   0.000     .5027662    .5900119
         10  |   .6646261   .0147728    44.99   0.000     .6356719    .6935803
         11  |   .7652831   .0132203    57.89   0.000     .7393717    .7911945
         12  |   .8428718   .0139768    60.31   0.000     .8154779    .8702658
```

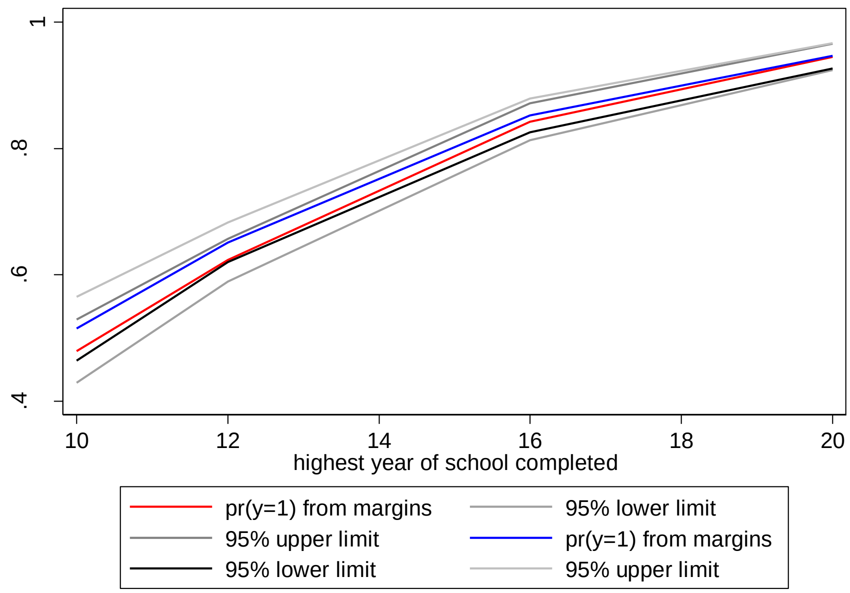
```
        13  |   .8982235   .0134213    66.93   0.000     .8719183    .9245288
        14  |    .935567    .011543    81.05   0.000     .9129432    .9581908
------------------------------------------------------------------------------
```

```
. marginsplot

  Variables that uniquely identify margins: age sex
```

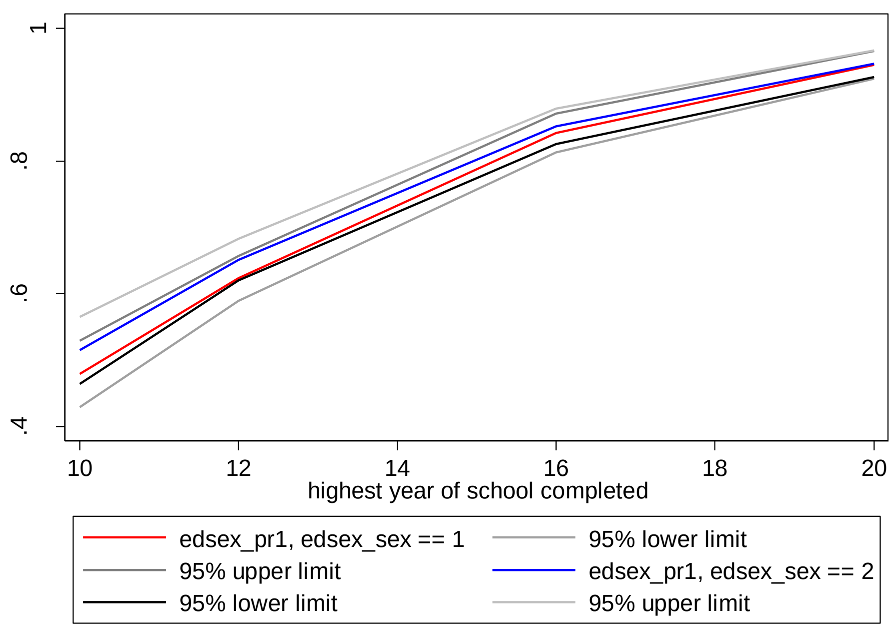



If you want to be able to format these graphs in your own ways, you can save predictions from margins into variables using mgen command:

```
. mgen, at(educ=(10 12 16 20) sex=(1 2 ) ) atmeans stub(edsex_)

Predictions from: margins, at(educ=(10 12 16 20) sex=(1 2)) atmeans predict(pr)

Variable    Obs Unique      Mean       Min       Max  Label
---------------------------------------------------------------------------------
edsex_pr1     8      8  .7320941  .4793427  .9470915  pr(y=1) from margins
edsex_ll1     8      8   .699113   .429009  .9269149  95% lower limit
edsex_ul1     8      8  .7650752  .5296764  .9672681  95% upper limit
edsex_educ    8      4      14.5        10        20  highest year of school co...
edsex_sex     8      2       1.5         1         2  respondents sex
---------------------------------------------------------------------------------

Specified values of covariates

                   2.          2.          3.          4.          5.
      age        born     marital     marital     marital     marital      childs
---------------------------------------------------------------------------
 46.93591    .0644788    .0926641    .1617761    .0351351    .2428571    1.838996

. graph twoway (line edsex_pr1 edsex_ll1 edsex_ul1 edsex_educ if edsex_sex==1,
>  sort lcolor(red))  (line edsex_pr1 edsex_ll1 edsex_ul1  edsex_educ if edsex
> _sex==2, sort lcolor(blue))
```




```
. separate edsex_pr1, by(edsex_sex)

              storage   display    value
variable name   type    format     label      variable label
------------------------------------------------------------------------------
edsex_pr11      float   %9.0g                 edsex_pr1, edsex_sex == 1
edsex_pr12      float   %9.0g                 edsex_pr1, edsex_sex == 2

. graph twoway (line edsex_pr11 edsex_ll1 edsex_ul1 edsex_educ if edsex_sex==1
> , sort lcolor(red))  (line edsex_pr12 edsex_ll1 edsex_ul1  edsex_educ if eds
> ex_sex==2, sort lcolor(blue))
```



3. Changes in Predicted Probabilities

Another way to interpret logistic regression results is using changes in predicted probabilities. These are changes in probability of the outcome as one variable changes, holding all other variables constant at certain values. There are two ways to measure such changes – discrete change and marginal effect.

*A. Discrete change*

Discrete change is a change in predicted probabilities corresponding to a given change in the independent variable. To obtain these, we calculate two probabilities and then calculate the difference between them. For example:

```
. mtable, at(sex=1) atmeans rowname(sex=1) statistics(ci)

Expression: Pr(vote), predict()

          |    Pr(y)        ll        ul
----------+-----------------------------
    sex=1 |    0.713     0.684     0.742

Specified values of covariates

          |                             2.           2.          3.          4.
          |      age       sex        born      marital     marital     marital
----------+-------------------------------------------------------------------
  Current |     46.9         1       .0645        .0927        .162       .0351

          |        5.
          |   marital    childs       educ
----------+------------------------------
  Current |      .243      1.84       13.4

. mtable, at(sex=2) atmeans rowname(sex=2) statistics(ci) below

Expression: Pr(vote), predict()

          |    Pr(y)        ll        ul
----------+-----------------------------
    sex=1 |    0.713     0.684     0.742
    sex=2 |    0.735     0.710     0.761

Specified values of covariates

          |                             2.           2.          3.          4.
          |      age       sex        born      marital     marital     marital
----------+-------------------------------------------------------------------
    Set 1 |     46.9         1       .0645        .0927        .162       .0351
  Current |     46.9         2       .0645        .0927        .162       .0351

          |        5.
          |   marital    childs       educ
----------+------------------------------
    Set 1 |      .243      1.84       13.4
  Current |      .243      1.84       13.4

. mtable, dydx(sex) atmeans rowname(sex=2 - sex=1) statistics(ci) below brief

Expression: Pr(vote), predict()

               |    Pr(y)        ll        ul
---------------+-----------------------------
         sex=1 |    0.713     0.684     0.742
```

```
         sex=2 |     0.735     0.710     0.761
 sex=2 - sex=1 |     0.022    -0.016     0.060
```

We can also calculate a bunch of predictions and then conduct pairwise comparisons and get significance tests for them using mlincom (need post option in mtable):

```
. mtable, at(sex=(1 2) marital=(1(1)5)) atmeans post
Expression: Pr(vote), predict()

           |      sex   marital     Pr(y)
 ----------+-----------------------------
         1 |        1         1     0.763
         2 |        1         2     0.660
         3 |        1         3     0.639
         4 |        1         4     0.692
         5 |        1         5     0.677
         6 |        2         1     0.783
         7 |        2         2     0.684
         8 |        2         3     0.665
         9 |        2         4     0.715
        10 |        2         5     0.701
Specified values of covariates
           |                 2.
           |      age      born    childs      educ
 ----------+---------------------------------------
   Current |     46.9     .0645      1.84      13.4

. mat list e(b)
e(b)[1,10]
            1.          2.          3.          4.          5.          6.
          _at         _at         _at         _at         _at         _at
y1  .76342597  .65995848  .63936008  .69224379  .67726949   .7829323

            7.          8.          9.         10.
          _at         _at         _at         _at
y1  .68447002  .66460183  .71543157  .70109834

. mlincom 1 - 6
             |   lincom    pvalue        ll        ul
-------------+----------------------------------------
           1 |   -0.020     0.250    -0.053     0.014
```

But there are commands that make it easier to do.

```
. logit vote age i.sex i.born i.marital childs educ
Iteration 0:   log likelihood = -1616.8899
Iteration 1:   log likelihood = -1361.6039
Iteration 2:   log likelihood = -1352.4837
Iteration 3:   log likelihood = -1352.4548
Iteration 4:   log likelihood = -1352.4548
Logistic regression                               Number of obs   =       2590
                                                  LR chi2(9)      =     528.87
                                                  Prob > chi2     =     0.0000
Log likelihood = -1352.4548                       Pseudo R2       =     0.1635
------------------------------------------------------------------------------
        vote |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0476294    .003864    12.33   0.000     .0400561    .0552027
             |
         sex |
     female  |   .1112819   .0966378     1.15   0.250    -.0781248    .3006886
             |
```

```
        born |
          no  |  -.9778304   .1865018    -5.24   0.000    -1.343367   -.6122936
             |
     marital |
     widowed  |  -.5084458   .2090768    -2.43   0.015    -.9182288   -.0986628
    divorced  |  -.5989672   .1349731    -4.44   0.000    -.8635096   -.3344249
   separated  |  -.3609247   .2461983    -1.47   0.143    -.8434645     .121615
never mar..  |  -.4303034   .1293215    -3.33   0.001    -.6837689   -.1768379
             |
      childs |  -.0350106   .0336983    -1.04   0.299     -.101058    .0310368
        educ |   .2879809   .0198907    14.48   0.000     .2489958    .3269661
       _cons |  -4.793928   .3483981   -13.76   0.000    -5.476775    -4.11108
------------------------------------------------------------------------------

. mchange
logit: Changes in Pr(y) | Number of obs = 2590
Expression: Pr(vote), predict(pr)
                            |    Change    p-value
----------------------------+---------------------
age                         |
                         +1 |     0.008      0.000
                        +SD |     0.125      0.000
                   Marginal |     0.008      0.000
sex                         |
             female vs male |     0.019      0.250
born                        |
                  no vs yes |    -0.185      0.000
marital                     |
         widowed vs married |    -0.089      0.019
        divorced vs married |    -0.106      0.000
       separated vs married |    -0.062      0.160
   never married vs married |    -0.075      0.001
        divorced vs widowed |    -0.017      0.682
       separated vs widowed |     0.027      0.626
   never married vs widowed |     0.014      0.746
      separated vs divorced |     0.044      0.355
  never married vs divorced |     0.031      0.278
 never married vs separated |    -0.013      0.788
childs                      |
                         +1 |    -0.006      0.301
                        +SD |    -0.010      0.302
                   Marginal |    -0.006      0.298
educ                        |
                         +1 |     0.048      0.000
                        +SD |     0.128      0.000
                   Marginal |     0.050      0.000

Average predictions
             |         0          1
-------------+----------------------
  Pr(y|base) |     0.317      0.683
```

Here we can see how probability changes when we go up by 1 unit (on average) and when we go up by 1 SD. For dichotomies, it is the difference between two categories. If values of independent variables are specified, predictions are computed at these values. For variables whose values are not specificed, changes are averaged across observed values (i.e., margins' asobserved option). Compare:

```
. mchange, atmeans
logit: Changes in Pr(y) | Number of obs = 2590
Expression: Pr(vote), predict(pr)
```

```
                           |    Change    p-value
---------------------------+----------------------
age                        |
                        +1 |     0.009      0.000
                       +SD |     0.132      0.000
                  Marginal |     0.009      0.000
sex                        |
            female vs male |     0.022      0.251
born                       |
                 no vs yes |    -0.224      0.000
marital                    |
         widowed vs married |    -0.101      0.024
        divorced vs married |    -0.121      0.000
       separated vs married |    -0.069      0.171
   never married vs married |    -0.084      0.001
        divorced vs widowed |    -0.020      0.680
       separated vs widowed |     0.032      0.626
   never married vs widowed |     0.017      0.747
      separated vs divorced |     0.052      0.350
  never married vs divorced |     0.037      0.280
 never married vs separated |    -0.015      0.787
childs                     |
                        +1 |    -0.007      0.303
                       +SD |    -0.012      0.305
                  Marginal |    -0.007      0.299
educ                       |
                        +1 |     0.054      0.000
                       +SD |     0.134      0.000
                  Marginal |     0.057      0.000

Predictions at base value
             |        0          1
-------------+----------------------
  Pr(y|base) |     0.274      0.726

Base values of regressors
             |                    2.          2.          2.          3.          4.
             |       age         sex        born     marital     marital     marital
-------------+-------------------------------------------------------------------------
          at |      46.9        .553       .0645       .0927        .162       .0351

             |         5.
             |   marital     childs       educ
-------------+---------------------------------
          at |      .243       1.84       13.4

1: Estimates with margins option atmeans.
```

We can also request more change units by using amount option or delta option, as well as more stats; we can also limit this investigation to certain variables:

```
. mchange, amount(all)
logit: Changes in Pr(y) | Number of obs = 2590
Expression: Pr(vote), predict(pr)
                           |    Change    p-value
---------------------------+----------------------
age                        |
                    0 to 1 |     0.008      0.000
                        +1 |     0.008      0.000
                       +SD |     0.125      0.000
                     Range |     0.505      0.000
                  Marginal |     0.008      0.000
```

```
sex                          |
             female vs male |     0.019      0.250
born                         |
                  no vs yes |    -0.185      0.000
marital                      |
         widowed vs married |    -0.089      0.019
        divorced vs married |    -0.106      0.000
       separated vs married |    -0.062      0.160
   never married vs married |    -0.075      0.001
        divorced vs widowed |    -0.017      0.682
       separated vs widowed |     0.027      0.626
   never married vs widowed |     0.014      0.746
      separated vs divorced |     0.044      0.355
  never married vs divorced |     0.031      0.278
 never married vs separated |    -0.013      0.788
childs                       |
                     0 to 1 |    -0.006      0.291
                         +1 |    -0.006      0.301
                        +SD |    -0.010      0.302
                      Range |    -0.050      0.305
                   Marginal |    -0.006      0.298
educ                         |
                     0 to 1 |     0.020      0.000
                         +1 |     0.048      0.000
                        +SD |     0.128      0.000
                      Range |     0.858      0.000
                   Marginal |     0.050      0.000

Average predictions
             |         0          1
-------------+----------------------
  Pr(y|base) |     0.317      0.683


. mchange educ, delta(5) statistics(all)
logit: Changes in Pr(y) | Number of obs = 2590
Expression: Pr(vote), predict(pr)
             |    Change    p-value         LL         UL    z-value
-------------+-------------------------------------------------------
educ         |
          +1 |     0.048      0.000      0.043      0.054     17.498
      +delta |     0.195      0.000      0.177      0.212     22.334
    Marginal |     0.050      0.000      0.044      0.056     16.917

             |   Std Err       From         To
-------------+---------------------------------
educ         |
          +1 |     0.003      0.683      0.732
      +delta |     0.009      0.683      0.878
    Marginal |     0.003         .z         .z

Average predictions
             |         0          1
-------------+----------------------
  Pr(y|base) |     0.317      0.683

1: Delta equals 5.
```

We can get these changes for a more limited range than min to max:

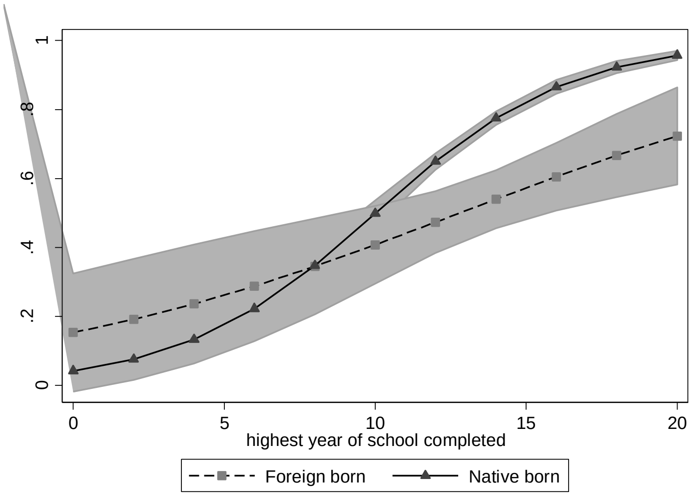
```
. mchange, amount(range) trim(5)
logit: Changes in Pr(y) | Number of obs = 2590
Expression: Pr(vote), predict(pr)
                            |    Change    p-value
----------------------------+----------------------
age                         |
                 5% to 95% |     0.428      0.000
sex                         |
            female vs male |     0.019      0.250
born                        |
                 no vs yes |    -0.185      0.000
marital                     |
        widowed vs married |    -0.089      0.019
       divorced vs married |    -0.106      0.000
      separated vs married |    -0.062      0.160
  never married vs married |    -0.075      0.001
       divorced vs widowed |    -0.017      0.682
      separated vs widowed |     0.027      0.626
  never married vs widowed |     0.014      0.746
     separated vs divorced |     0.044      0.355
 never married vs divorced |     0.031      0.278
never married vs separated |    -0.013      0.788
childs                      |
                 5% to 95% |    -0.031      0.300
educ                        |
                 5% to 95% |     0.448      0.000

Average predictions
             |        0          1
-------------+----------------------
  Pr(y|base) |     0.317      0.683


. centile educ, centile(0 5 95 100)
                                                       -- Binom. Interp. --
    Variable |     Obs  Percentile      Centile        [95% Conf. Interval]
-------------+-------------------------------------------------------------
        educ |    2753          0             0               0           0*
             |                  5             8               8           9
             |                 95            18              18          18
             |                100            20              20          20*
* Lower (upper) confidence limit held at minimum (maximum) of sample
```

People often conclude that two groups are different if confidence intervals do not overlap – but that is usually too conservative. Looking at discrete changes with a confidence interval is more informative. Note that if you have linked variables – variables with squared or cubed terms, or with interactions – you should use factor variable notation (as in the example below), and then the commands will keep track of that for you when generating predictions.

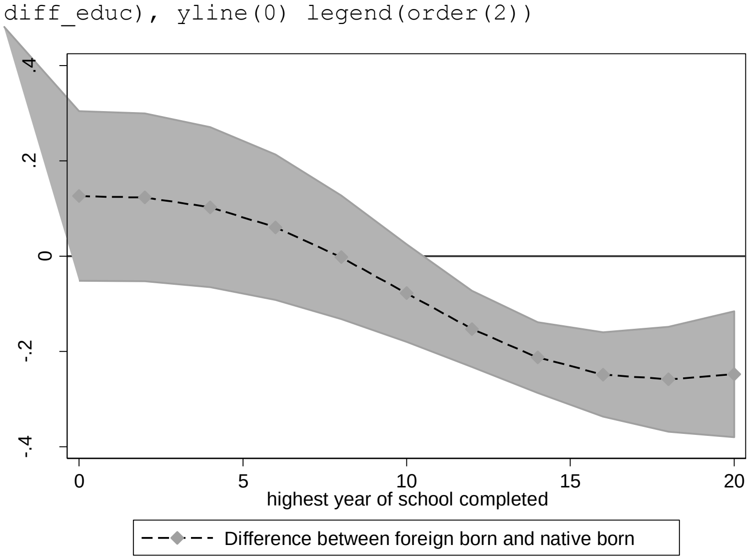
```
. logit vote childs i.sex i.born##c.educ i.marital  age
Iteration 0:   log likelihood = -1616.8899
Iteration 1:   log likelihood = -1357.6437
Iteration 2:   log likelihood = -1347.2285
Iteration 3:   log likelihood = -1347.1953
Iteration 4:   log likelihood = -1347.1953
Logistic regression                               Number of obs   =       2590
                                                  LR chi2(10)     =     539.39
                                                  Prob > chi2     =     0.0000
Log likelihood = -1347.1953                       Pseudo R2       =     0.1668
-------------------------------------------------------------------------------
```

```
          vote |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------+----------------------------------------------------------------
        childs |  -.0340252   .0337392    -1.01   0.313    -.1001528    .0321024
               |
           sex |
       female  |   .1040851   .0968511     1.07   0.283    -.0857396    .2939098
               |
          born |
           no  |    1.40973   .7190502     1.96   0.050      .000417    2.819042
          educ |    .310949   .0214335    14.51   0.000       .26894     .352958
               |
   born#c.educ |
           no  |   -.177672   .0521626    -3.41   0.001    -.2799088   -.0754351
               |
       marital |
      widowed  |  -.4717709   .2097889    -2.25   0.025    -.8829496   -.0605922
     divorced  |  -.5976599   .1353399    -4.42   0.000    -.8629213   -.3323985
    separated  |  -.3473378   .2469307    -1.41   0.160    -.8313131    .1366374
never married  |  -.4409393   .1298164    -3.40   0.001    -.6953748   -.1865039
               |
           age |   .0476348   .0038642    12.33   0.000     .0400611    .0552086
         _cons |  -5.086571   .3639937   -13.97   0.000    -5.799985   -4.373156
--------------------------------------------------------------------------------

. mgen, atmeans at(educ=(0(2)20) born=1) stub(nb_)
Predictions from: margins, atmeans at(educ=(0(2)20) born=1) predict(pr)
Variable   Obs Unique      Mean       Min       Max  Label
------------------------------------------------------------------------------------
nb_pr1      11     11  .4991813  .0424739  .9570376  pr(y=1) from margins
nb_ll1      11     11  .4690912  .0204639  .9437191  95% lower limit
nb_ul1      11     11  .5292713  .0644839   .970356  95% upper limit
nb_educ     11     11        10         0        20  highest year of school com...
------------------------------------------------------------------------------------

Specified values of covariates
                 2.                      2.         3.         4.         5.
   childs        sex       born    marital    marital    marital    marital
------------------------------------------------------------------------------
 1.838996   .5532819          1   .0926641   .1617761   .0351351   .2428571

       age
-----------
  46.93591

. mgen, atmeans at(educ=(0(2)20) born=2) stub(fb_)
Predictions from: margins, atmeans at(educ=(0(2)20) born=2) predict(pr)
Variable   Obs Unique      Mean        Min       Max  Label
------------------------------------------------------------------------------------
fb_pr1      11     11  .4209988   .1537178  .7230831  pr(y=1) from margins
fb_ll1      11     11  .2880444  -.0171945  .5822673  95% lower limit
fb_ul1      11     11  .5539532   .3246301  .8638988  95% upper limit
fb_educ     11     11        10          0        20  highest year of school co...
------------------------------------------------------------------------------------

Specified values of covariates
                 2.                      2.         3.         4.         5.
   childs        sex       born    marital    marital    marital    marital
------------------------------------------------------------------------------
 1.838996   .5532819          2   .0926641   .1617761   .0351351   .2428571

       age
-----------
  46.93591
```

```
. lab var fb_pr1 "Foreign born"
. lab var nb_pr1 "Native born"
. graph twoway (rarea nb_ul1 nb_ll1 nb_educ, col(gs10)) (rarea fb_ul1 fb_ll1 fb_educ,
color(gs10)) (connected  fb_pr1 nb_pr1 fb_educ, lpattern(dash solid)), legend(order(3
4))
```




```
. mgen, dydx(born) at(educ=(0(2)20)) stub(diff_)

Predictions from: margins, dydx(born) at(educ=(0(2)20)) predict(pr)

Variable    Obs Unique       Mean        Min        Max  Label
---------------------------------------------------------------------------------------
diff_d_pr1   11     11  -.0717905  -.2587097   .1261362  d_pr(y=1) from margins
diff_ll1     11     11  -.1983488  -.3805256  -.0519222  95% lower limit
diff_ul1     11     11   .0547677  -.1604816   .3041946  95% upper limit
diff_educ    11     11         10          0         20  highest year of school
completed
---------------------------------------------------------------------------------------

. lab var diff_d_pr1 "Difference between foreign born and native born"

. graph twoway (rarea diff_ul1 diff_ll1 diff_educ, col(gs10)) (connected  diff_d_pr1
diff_educ), yline(0) legend(order(2))
```



*B. Marginal effects.*

One thing that we saw in the mchange output above but did not discuss yet is marginal effects – these are partial derivatives, slopes of probability curve at a certain set of values of independent variables. Marginal effects, of course, vary along X; they are the largest at the value of X that corresponds to P(Y=1|X)=.5 – this can be seen in the graph.

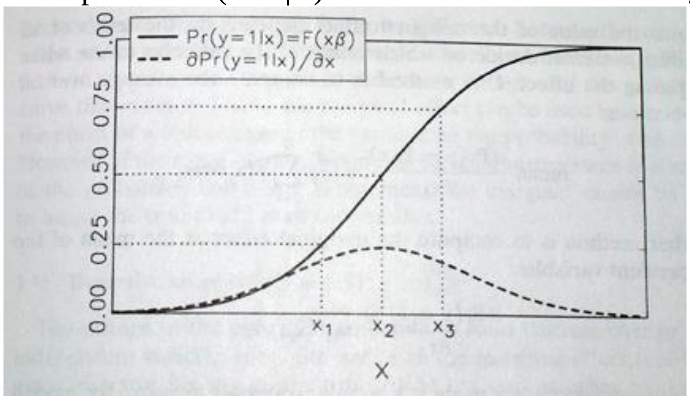



**Figure 3.12.** Marginal Effect in the Binary Response Model

The following graph compares a marginal change and a discrete change at a specific point:

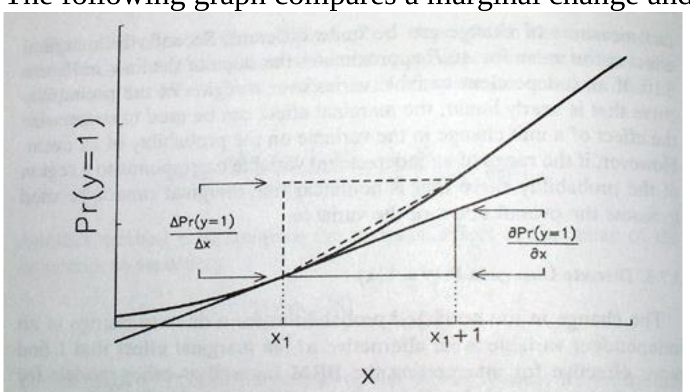



**Figure 3.13.** Partial Change Versus Discrete Change in Nonlinear Models

Marginal effects are inappropriate for binary independent variables; that's why discrete changes are reported for those instead.

There are three ways that marginal effects are usually estimated:

1. Marginal effects at the mean (MEM)
2. Marginal effects at representative values (MER)
3. Average marginal effects (AME) (marginal effects are estimated at all values and then averaged out)

```
. logit vote age i.sex i.born i.marital childs educ

Iteration 0:   log likelihood = -1616.8899
Iteration 1:   log likelihood = -1361.6039
Iteration 2:   log likelihood = -1352.4837
Iteration 3:   log likelihood = -1352.4548
Iteration 4:   log likelihood = -1352.4548

Logistic regression                               Number of obs   =       2590
```

```
                                                  LR chi2(9)      =     528.87
                                                  Prob > chi2     =     0.0000
Log likelihood = -1352.4548                       Pseudo R2       =     0.1635

------------------------------------------------------------------------------
        vote |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0476294    .003864    12.33   0.000     .0400561    .0552027
             |
         sex |
     female  |   .1112819   .0966378     1.15   0.250    -.0781248    .3006886
             |
        born |
         no  |  -.9778304   .1865018    -5.24   0.000    -1.343367   -.6122936
             |
     marital |
    widowed  |  -.5084458   .2090768    -2.43   0.015    -.9182288   -.0986628
   divorced  |  -.5989672   .1349731    -4.44   0.000    -.8635096   -.3344249
  separated  |  -.3609247   .2461983    -1.47   0.143    -.8434645     .121615
never mar..  |  -.4303034   .1293215    -3.33   0.001    -.6837689   -.1768379
             |
      childs |  -.0350106   .0336983    -1.04   0.299     -.101058    .0310368
        educ |   .2879809   .0198907    14.48   0.000     .2489958    .3269661
       _cons |  -4.793928   .3483981   -13.76   0.000    -5.476775    -4.11108
------------------------------------------------------------------------------
```

<u>Average marginal effects (AME):</u>

```
. mchange

logit: Changes in Pr(y) | Number of obs = 2590

Expression: Pr(vote), predict(pr)

                            |    Change    p-value
----------------------------+----------------------
age                         |
                         +1 |     0.008      0.000
                        +SD |     0.125      0.000
                   Marginal |     0.008      0.000
sex                         |
             female vs male |     0.019      0.250
born                        |
                  no vs yes |    -0.185      0.000
marital                     |
         widowed vs married |    -0.089      0.019
        divorced vs married |    -0.106      0.000
       separated vs married |    -0.062      0.160
   never married vs married |    -0.075      0.001
        divorced vs widowed |    -0.017      0.682
       separated vs widowed |     0.027      0.626
   never married vs widowed |     0.014      0.746
      separated vs divorced |     0.044      0.355
  never married vs divorced |     0.031      0.278
 never married vs separated |    -0.013      0.788
childs                      |
                         +1 |    -0.006      0.301
                        +SD |    -0.010      0.302
                   Marginal |    -0.006      0.298
educ                        |
                         +1 |     0.048      0.000
                        +SD |     0.128      0.000
                   Marginal |     0.050      0.000
```

```
Average predictions

             |         0          1
-------------+----------------------
  Pr(y|base) |     0.317      0.683
```

In addition to mchange, we can also obtain marginal effects with dydx option in margins:

```
. margins, dydx(*)

Average marginal effects                          Number of obs   =       2590
Model VCE    : OIM

Expression   : Pr(vote), predict()
dy/dx w.r.t. : age 2.sex 2.born 2.marital 3.marital 4.marital 5.marital
               childs educ

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0083074   .0006053    13.72   0.000      .007121    .0094937
             |
         sex |
     female  |   .0194592    .016928     1.15   0.250    -.0137191    .0526375
             |
        born |
         no  |  -.1851289   .0364786    -5.07   0.000    -.2566257   -.1136321
             |
     marital |
    widowed  |  -.0892473   .0380707    -2.34   0.019    -.1638646   -.0146301
   divorced  |  -.1062677   .0244728    -4.34   0.000    -.1542335   -.0583019
  separated  |  -.0621571    .044188    -1.41   0.160     -.148764    .0244498
never mar..  |  -.0747909   .0231535    -3.23   0.001    -.1201708   -.0294109
             |
      childs |  -.0061064   .0058731    -1.04   0.298    -.0176175    .0054047
        educ |   .0502287   .0029691    16.92   0.000     .0444093    .0560481
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

<u>Marginal effects at the mean (MEM):</u>

```
. mchange, atmeans

logit: Changes in Pr(y) | Number of obs = 2590

Expression: Pr(vote), predict(pr)

                            |    Change    p-value
----------------------------+----------------------
age                         |
                         +1 |     0.009      0.000
                        +SD |     0.132      0.000
                   Marginal |     0.009      0.000
sex                         |
             female vs male |     0.022      0.251
born                        |
                  no vs yes |    -0.224      0.000
marital                     |
         widowed vs married |    -0.101      0.024
```

```
        divorced vs married |    -0.121      0.000
       separated vs married |    -0.069      0.171
   never married vs married |    -0.084      0.001
        divorced vs widowed |    -0.020      0.680
       separated vs widowed |     0.032      0.626
   never married vs widowed |     0.017      0.747
      separated vs divorced |     0.052      0.350
  never married vs divorced |     0.037      0.280
 never married vs separated |    -0.015      0.787
childs                      |
                         +1 |    -0.007      0.303
                        +SD |    -0.012      0.305
                   Marginal |    -0.007      0.299
educ                        |
                         +1 |     0.054      0.000
                        +SD |     0.134      0.000
                   Marginal |     0.057      0.000

Predictions at base value

             |         0          1
-------------+----------------------
  Pr(y|base) |     0.274      0.726

Base values of regressors

             |                    2.          2.          2.          3.          4.
             |       age         sex        born     marital     marital     marital
-------------+------------------------------------------------------------------------
          at |      46.9        .553       .0645       .0927        .162       .0351

             |         5.
             |   marital      childs        educ
-------------+------------------------------------
          at |      .243        1.84        13.4

1: Estimates with margins option atmeans.
```

We can also get them centered at means (the default option shows mean+1):

```
. mchange, atmeans centered

logit: Changes in Pr(y) | Number of obs = 2590

Expression: Pr(vote), predict(pr)

                            |    Change    p-value
----------------------------+----------------------
age                         |
                +1 centered |     0.009      0.000
               +SD centered |     0.162      0.000
                   Marginal |     0.009      0.000
sex                         |
             female vs male |     0.022      0.251
born                        |
                  no vs yes |    -0.224      0.000
marital                     |
         widowed vs married |    -0.101      0.024
        divorced vs married |    -0.121      0.000
       separated vs married |    -0.069      0.171
   never married vs married |    -0.084      0.001
        divorced vs widowed |    -0.020      0.680
```

```
       separated vs widowed |     0.032      0.626
   never married vs widowed |     0.017      0.747
      separated vs divorced |     0.052      0.350
  never married vs divorced |     0.037      0.280
 never married vs separated |    -0.015      0.787
childs                      |
               +1 centered |    -0.007      0.299
               +SD centered |    -0.012      0.299
                  Marginal |    -0.007      0.299
educ                        |
               +1 centered |     0.057      0.000
               +SD centered |     0.167      0.000
                  Marginal |     0.057      0.000

Predictions at base value

             |         0          1
-------------+----------------------
  Pr(y|base) |     0.274      0.726

Base values of regressors

             |                    2.          2.          2.          3.          4.
             |       age         sex        born     marital     marital     marital
-------------+----------------------------------------------------------------------
          at |      46.9        .553       .0645       .0927        .162       .0351

             |        5.
             |   marital      childs        educ
-------------+----------------------------------
          at |      .243        1.84        13.4

1: Estimates with margins option atmeans.
```

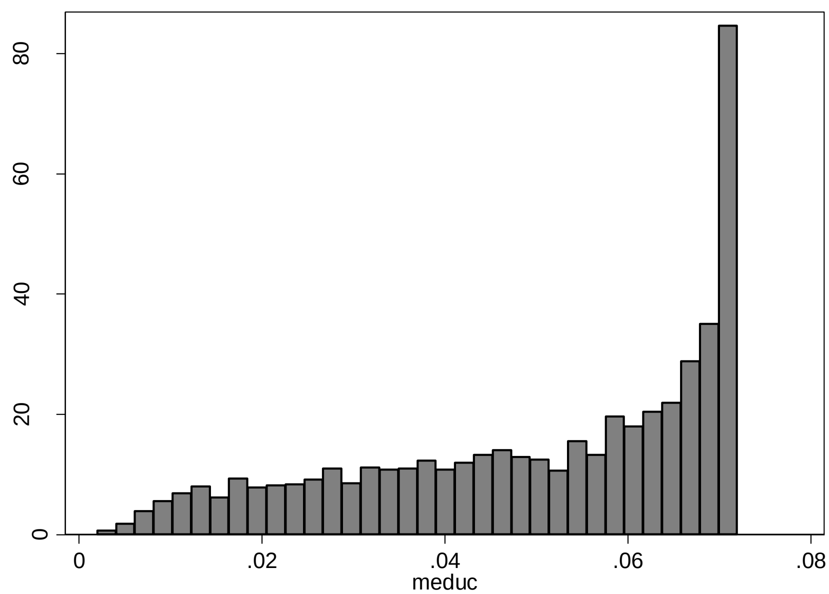
In case of logistic regression, marginal effect for X can be calculated as P(Y=1|X)*P(Y=0|X)*b; For example, we can replicate the result for MEM:

```
. margins, atmeans

Adjusted predictions                              Number of obs   =       2590
Model VCE    : OIM

Expression   : Pr(vote), predict()
at           : age             =    46.93591 (mean)
               1.sex           =    .4467181 (mean)
               2.sex           =    .5532819 (mean)
               1.born          =    .9355212 (mean)
               2.born          =    .0644788 (mean)
               1.marital       =    .4675676 (mean)
               2.marital       =    .0926641 (mean)
               3.marital       =    .1617761 (mean)
               4.marital       =    .0351351 (mean)
               5.marital       =    .2428571 (mean)
               childs          =    1.838996 (mean)
               educ            =    13.39459 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .7255038   .0100492    72.20   0.000     .7058078    .7451997
------------------------------------------------------------------------------
```

```
. di .7255038*(1-.7255038)* .2879809
.05735083
```

Histogram of marginal effects can help us better understand whether MEM or AME better represent what is going on in our sample:

```
. predict double prhat if e(sample)
(option pr assumed; Pr(vote))
(175 missing values generated)

. gen double meduc=prhat*(1-prhat) *_b[educ]
(175 missing values generated)

. histogram meduc
(bin=34, start=.00199118, width=.00205894)
```




<u>Marginal effects at representative values (MER):</u>

```
. mchange educ, at(educ=12)

logit: Changes in Pr(y) | Number of obs = 2590

Expression: Pr(vote), predict(pr)

             |    Change    p-value
-------------+----------------------
educ         |
          +1 |     0.057      0.000
         +SD |     0.154      0.000
    Marginal |     0.059      0.000

Average predictions

             |         0          1
-------------+----------------------
  Pr(y|base) |     0.382      0.618

Base values of regressors
```

```
             |      educ
-------------+-----------
          at |        12


. mchange educ, at(educ=16)

logit: Changes in Pr(y) | Number of obs = 2590

Expression: Pr(vote), predict(pr)

             |    Change    p-value
-------------+----------------------
educ         |
          +1 |     0.036      0.000
         +SD |     0.090      0.000
    Marginal |     0.039      0.000

Average predictions

             |         0          1
-------------+----------------------
  Pr(y|base) |     0.182      0.818

Base values of regressors

             |      educ
-------------+-----------
          at |        16


. mchange educ, at(educ=10)

logit: Changes in Pr(y) | Number of obs = 2590

Expression: Pr(vote), predict(pr)

             |    Change    p-value
-------------+----------------------
educ         |
          +1 |     0.061      0.000
         +SD |     0.174      0.000
    Marginal |     0.061      0.000

Average predictions

             |         0          1
-------------+----------------------
  Pr(y|base) |     0.503      0.497

Base values of regressors

             |      educ
-------------+-----------
          at |        10
```