Marginal Effects & Discrete Change

[Notes adapted from Long 1997 and Long and Freese 2003. See also http://www.ugrad.math.ubc.ca/coursedoc/math100/notes/derivative/slope.html. My calculus is a bit weak after 30 years of little use so corrections are welcome.]

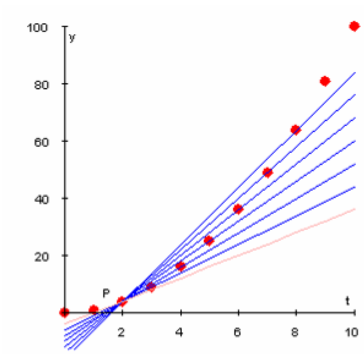
Note: The user-written `mfx2` and `margeff` commands are often attractive alternatives to Stata's own `mfx` command. Both are good for automating tasks that could otherwise take several commands and for reformatting output so it can easily be placed into tables. There are also several other user-written routines that can be good alternatives to `mfx` under certain circumstances. We'll talk about some of these commands later. Also, in Stata 11, the new `margins` command supercedes `mfx`, but `mfx` still works.

Definition. In binary regression models, the marginal effect is the slope of the probability curve relating X_k to $Pr(Y=1|X)$, holding all other variables constant. A little calculus review will help make this clearer.

Simple Explanation. Draw a graph of $F(X_k)$ against X_k, holding all the other X's constant (e.g. at their means). Chose 2 points, $[X_k, F(X_k)]$ and $[X_k+\delta, F(X_k+\delta)$. When δ is very very small, the slope of the line connecting the two points will equal or almost equal the marginal effect of X_k.

More Detailed Explanation. What is the slope of a curve? Intuitively, think of it this way. Draw a graph of F(X) against X, e.g. $F(X) = X^2$. Chose specific values of X and F(X), e.g. [2, F(2)]. Choose another point, e.g. [8, F(8)]. Draw a line connecting the points. This line has a slope. The slope is the *average rate of change*.

Now, choose another point that is closer to [2, F(2)], e.g. [7, F(7)]. Draw a line connecting these points. This too will have a slope. Keep on choosing points that are closer to [2, F(2)]. The *instantaneous rate of change* is the limit of the slopes for the lines connecting [X, F(X)] and $[X+\delta, F(X+\delta)]$ as δ gets closer and closer to 0.



Source: http://www.ugrad.math.ubc.ca/coursedoc/math100/notes/derivative/slope.html
See the above link for more information

Calculus is used to compute slopes (& marginal effects). For example, if $Y = X^2$, then the slope is 2X. Hence, if X = 2, the slope is 4. The following table illustrates this. Note that, as δ gets smaller and smaller, the slope gets closer and closer to 4.

F(X) = X^2   X = 2   F(2) = 4

δ	X+δ	F(X+δ)	Change in F(X)	Change in X	Slope
100	102	10404	10400	100	104
10	12	144	140	10	14
1	3	9	5	1	5
0.1	2.1	4.41	0.41	0.1	4.1
0.01	2.01	4.0401	0.0401	0.01	4.01
0.001	2.001	4.004001	0.004001	0.001	4.001
0.0001	2.0001	4.00040001	0.00040001	0.0001	4.0001

Marginal effects are also instantaneous rates of change; you compute them for a variable while all other variables are held constant. The magnitude of the marginal effect depends on the values of the other variables and their coefficients. The *marginal effect at the mean* is popular (i.e. compute the marginal effects when all x's are at their mean) but other options are possible.

Logistic Regression. Again, calculus is used to compute the marginal effects. In the case of logistic regression, F(X) = P(Y=1|X), and

$$\text{Marginal Effect for } X_k = P(Y=1 |X) * P(Y = 0|X) * b_k.$$

Example:

```
. use http://www.nd.edu/~rwilliam/xsoc73994/statafiles/glm-logit.dta, clear
. logit grade  gpa tuce psi, nolog

Logit estimates                                   Number of obs   =         32
                                                  LR chi2(3)      =      15.40
                                                  Prob > chi2     =     0.0015
Log likelihood = -12.889633                       Pseudo R2       =     0.3740

------------------------------------------------------------------------------
       grade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gpa |   2.826113   1.262941     2.24   0.025     .3507938    5.301432
        tuce |   .0951577   .1415542     0.67   0.501    -.1822835    .3725988
         psi |   2.378688   1.064564     2.23   0.025       .29218    4.465195
       _cons |  -13.02135   4.931325    -2.64   0.008    -22.68657    -3.35613
------------------------------------------------------------------------------
```

```
. mfx

Marginal effects after logit
      y  = Pr(grade) (predict)
         =  .25282025
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
     gpa |   .5338589      .23704    2.25   0.024   .069273  .998445   3.11719
    tuce |   .0179755      .02624    0.69   0.493  -.033448  .069399   21.9375
     psi*|   .4564984      .18105    2.52   0.012    .10164  .811357     .4375
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1

. prchange

logit: Changes in Predicted Probabilities for grade

      min->max      0->1     -+1/2    -+sd/2  MargEfct
 gpa    0.7872    0.0008    0.5055    0.2466    0.5339
tuce    0.2824    0.0038    0.0180    0.0701    0.0180
 psi    0.4565    0.4565    0.4330    0.2246    0.4493

              0       1
Pr(y|x)  0.7472  0.2528

            gpa     tuce      psi
    x=  3.11719  21.9375    .4375
sd(x)=  .466713  3.90151  .504016
```

Looking specifically at GPA – we can tell from the output that, when all variables are at their means, pr(Y=1|X) = .2528, Pr(Y=0|X) = .7472, and b_{GPA} = 2.826113. The marginal effect of GPA is therefore

$$\text{Marginal Effect of GPA} = P(Y=1|X) * P(Y=0|X) * b_{GPA} = .2528 * .7472 * 2.826113 = .5339$$

The `mfx` command reports this in the column labeled dy/dx while `prchange` reports it in the column labeled MargEfct.

The following table again shows you that, in logistic regression, as the distance between two points gets smaller and smaller, i.e. as δ gets closer and closer to 0, the slope of the line connecting the points gets closer and closer to the marginal effect.

Logistic Regression

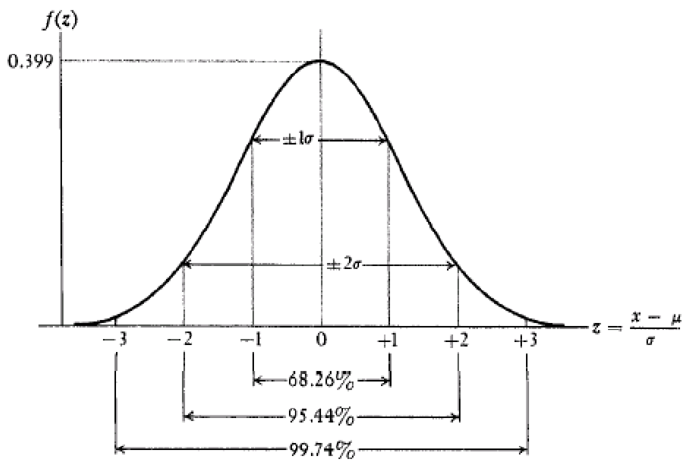
F(X,GPA) = P(Y=1|X, GPA) GPA=3.11719 Other X's at Mean F(X,3.11719) = .25282025107643

δ	GPA+δ	F(X,GPA+δ)	Change in F(X,GPA)	Change in GPA	Slope
10	13.11719	1	0.747179749	10	0.0747179749
1	4.11719	0.851002558	0.598182307	1	0.5981823074
0.1	3.21719	0.309808293	0.056988042	0.1	0.5698804165
0.05	3.16719	0.280431679	0.027611428	0.05	0.5522285640
0.01	3.12719	0.258196038	0.005375787	0.01	0.5375787403
0.001	3.11819	0.253354486	0.000534235	0.001	0.5342350788
0.0001	3.11729	0.252873644	5.3393E-05	0.0001	0.5339297264

Probit. In probit, the marginal effect is

$$\text{Marginal Effect for } X_k = \Phi(XB) * b_k$$

where Φ is the probability density function for a standardized normal variable. For example, as this diagram shows, Φ(0) = .399:



Example:

```
. use http://www.nd.edu/~rwilliam/xsoc73994/statafiles/glm-logit.dta, clear
. probit  grade gpa tuce psi, nolog

Probit estimates                                  Number of obs   =         32
                                                  LR chi2(3)      =      15.55
                                                  Prob > chi2     =     0.0014
Log likelihood = -12.818803                       Pseudo R2       =     0.3775

------------------------------------------------------------------------------
       grade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gpa |    1.62581   .6938818     2.34   0.019     .2658269    2.985794
        tuce |   .0517289   .0838901     0.62   0.537    -.1126927    .2161506
         psi |   1.426332    .595037     2.40   0.017     .2600814    2.592583
       _cons |   -7.45232   2.542467    -2.93   0.003    -12.43546   -2.469177
------------------------------------------------------------------------------

. mfx

Marginal effects after probit
      y  = Pr(grade) (predict)
         =  .26580809
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
     gpa |   .5333471      .23246    2.29   0.022   .077726  .988968   3.11719
    tuce |   .0169697      .02712    0.63   0.531  -.036184  .070123   21.9375
    psi*|    .464426      .17028    2.73   0.006   .130682   .79817     .4375
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

```
. prchange

probit: Changes in Predicted Probabilities for grade

       min->max      0->1     -+1/2    -+sd/2  MargEfct
 gpa    0.7814    0.0000    0.4992    0.2453    0.5333
tuce    0.2701    0.0046    0.0170    0.0661    0.0170
 psi    0.4644    0.4644    0.4446    0.2328    0.4679

              0       1
Pr(y|x)  0.7342  0.2658

             gpa     tuce      psi
    x=  3.11719  21.9375    .4375
sd(x)=  .466713  3.90151  .504016

. * marginal change for GPA.  The invnorm function gives us the Z-score for the stated
. * prob of success.  The normden function gives us the pdf value for that Z-score.
. display invnorm(.2658)
-.62556546
. display normden(invnorm(.2658))
.32804496
. display normden(invnorm(.2658)) * 1.62581
.53333878
```

Marginal Effect for GPA = $\Phi(XB) * b_k$ = .32804496 * 1.62581 = .5333

The following table again shows you that, in a probit model, as the distance between two points gets smaller and smaller, i.e. as δ gets closer and closer to 0, the slope of the line connecting the points gets closer and closer to the marginal effect.

Probit

F(X,GPA) = P(Y=1|X, GPA)   GPA=3.11719 Other X's at Mean   F(X,3.11719) = .26580811

δ	GPA+δ	F(X,GPA+δ)	Change in F(X,GPA)	Change in GPA	Slope
10	13.11719	1	0.73419189	10	0.0734191890
1	4.11719	0.841409951	0.575601841	1	0.5756018408
0.1	3.21719	0.321696605	0.055888495	0.1	0.5588849545
0.01	3.12719	0.27116854	0.00536043	0.01	0.5360430345
0.001	3.11819	0.266341711	0.000533601	0.001	0.5336010479
0.0001	3.11729	0.26586143	5.33203E-05	0.0001	0.5332029760

OLS. Here is what you get when you compute the marginal effects for OLS:

```
. use http://www.nd.edu/~rwilliam/xsoc73994/statafiles/glm-reg.dta, clear

. reg income  educ jobexp black

      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  3,   496) =  787.14
       Model |  33206.4588     3  11068.8196           Prob > F      =  0.0000
    Residual |  6974.79047   496  14.0620776           R-squared     =  0.8264
-------------+------------------------------           Adj R-squared =  0.8254
       Total |  40181.2493   499  80.5235456           Root MSE      =  3.7499

------------------------------------------------------------------------------
      income |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   1.840407   .0467507    39.37   0.000     1.748553    1.932261
      jobexp |   .6514259   .0350604    18.58   0.000     .5825406    .7203111
       black |   -2.55136   .4736266    -5.39   0.000    -3.481921   -1.620798
       _cons |   -4.72676   .9236842    -5.12   0.000    -6.541576   -2.911943
------------------------------------------------------------------------------

. mfx

Marginal effects after regress
      y  = Fitted values (predict)
         =      27.79
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
    educ |   1.840407      .04675   39.37   0.000   1.74878  1.93204     13.16
  jobexp |   .6514259      .03506   18.58   0.000   .582709  .720143     13.52
  black*|   -2.55136      .47363   -5.39   0.000  -3.47965 -1.62307         .2
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

The marginal effects are the same as the slope coefficients. This is because relationships are linear in OLS regression and do not vary depending on the values of the other variables.

Usefulness of Marginal Effects; Discrete Change as an Alternative. Marginal effects are popular in some disciplines (e.g. Economics) because they often provide a good approximation to the amount of change in Y that will be produced by a 1-unit change in X_k. Nonetheless, Long does not think they are very helpful for models with binary dependent variables. First, given the nonlinearity of the model, it is difficult to translate the marginal effects into the change that will occur if there is a discrete change (e.g. 1-unit change) in X_k. Second, marginal effects are inappropriate for binary independent variables. Long therefore prefers measures of discrete change. The discrete change for a change of δ in X_k (holding all other variables constant) is

$$Pr(Y = 1|X, X_k+\delta) - Pr(y=1|X, X_k)$$

For example, Long and Freese's `prchange` program shows how the probability of success changes as you go from the variable mean - .5 to the variable mean + .5, a 1-unit increase (holding all other variables constant). Example:

```
. quietly logit  grade gpa tuce psi
. prchange

logit: Changes in Predicted Probabilities for grade

      min->max      0->1     -+1/2    -+sd/2  MargEfct
 gpa    0.7872    0.0008    0.5055    0.2466    0.5339
tuce    0.2824    0.0038    0.0180    0.0701    0.0180
 psi    0.4565    0.4565    0.4330    0.2246    0.4493

              0       1
Pr(y|x)  0.7472  0.2528

            gpa     tuce      psi
    x=  3.11719  21.9375    .4375
sd(x)=  .466713  3.90151  .504016
```

The above shows us that the mean of gpa is 3.11719. The column labeled -+1/2 (referred to as the *centered discrete change*) shows us that, as gpa goes from 2.611719 to 3.611719 (holding all other variables at their means) the probability of success increases by .5055. By adding the `fromto` parameter we can get a little more detail:

```
. prchange, fromto

logit: Changes in Predicted Probabilities for grade

         from:       to:      dif:     from:       to:      dif:     from:       to:      dif:
        x=min     x=max  min->max       x=0       x=1      0->1     x-1/2     x+1/2     -+1/2
 gpa    0.0168    0.8040    0.7872    0.0001    0.0009    0.0008    0.0761    0.5816    0.5055
tuce    0.1162    0.3985    0.2824    0.0403    0.0441    0.0038    0.2439    0.2619    0.0180
 psi    0.1068    0.5633    0.4565    0.1068    0.5633    0.4565    0.0934    0.5264    0.4330

         from:       to:      dif:
      x-1/2sd   x+1/2sd    -+sd/2  MargEfct
 gpa    0.1489    0.3955    0.2466    0.5339
tuce    0.2194    0.2895    0.0701    0.0180
 psi    0.1567    0.3813    0.2246    0.4493

              0       1
Pr(y|x)  0.7472  0.2528

            gpa     tuce      psi
    x=  3.11719  21.9375    .4375
sd(x)=  .466713  3.90151  .504016
```

So, when GPA = 2.611719 and the other vars equal their means, the probability of success is .0761. When GPA = 3.611719, the probability of success is .5816, an increase of .5055.

You can use the `adjust` command to make sure `prchange` did it right:

```
. adjust gpa=2.61719 tuce psi, pr

-------------------------------------------------------------------
     Dependent variable: grade     Command: logit
 Covariates set to mean: tuce = 21.9375, psi = .4375
 Covariate set to value: gpa = 2.61719
-------------------------------------------------------------------

----------------------
      All |         pr
----------+-----------
          |    .076091
----------------------
     Key:  pr  =  Probability

. adjust gpa=3.61719 tuce psi, pr

-------------------------------------------------------------------
     Dependent variable: grade     Command: logit
 Covariates set to mean: tuce = 21.9375, psi = .4375
 Covariate set to value: gpa = 3.61719
-------------------------------------------------------------------

----------------------
      All |         pr
----------+-----------
          |    .581622
----------------------
     Key:  pr  =  Probability
```

`prchange` includes lots of other options, e.g. you can fix variables at values other than their mean.

As Long and Freese (2003) note (p. 142), the value of the discrete change depends on

- The start level of the variable that is being changed; e.g. do you want to examine the effect of age beginning at 30? Or 40? Or 50? Or at or about its mean?
- The amount of change in that variable. Do you want to examine the effect of a 1 year change, a 5 year change, or a 10 year change?
- The level of all other variables in the model. Do you want to hold all variables at their mean? Or, do you want to examine the effect for women? Or compute changes separately for men and women? The means are one logical choice, but you could choose other substantively interesting baselines. For example, if you were interested in the effects of education on labor force participation for young women without children, you could hold AGE at 30 and the # of children at 0, and hold all other variables at their means.

Discrete Change Versus Marginal Change. The discrete change and the marginal effects need not be equal. However, if the change in x_k occurs over a region of the probability curve that is roughly linear, the two measures will be close. When the changes in x_k are relatively small given the range of x_k, the marginal change and the discrete change can be very similar (conversely, they can be very dissimilar when the range of change is relatively large). We can see this by observing what happens as we change the way that GPA is scaled. In the following, a 1-unit change in GPA is initially very large relative to the scaling of GPA (in fact it is larger than the range of GPA) and then gradually gets smaller and smaller. Compare the columns labeled -+1/2 and MargEfct as the scaling of GPA changes.

```
. gen xgpa =gpa
. replace xgpa = gpa/10
. quietly logit grade xgpa tuce psi, nolog
. prchange xgpa

logit: Changes in Predicted Probabilities for grade

      min->max      0->1     -+1/2    -+sd/2  MargEfct
xgpa    0.7872    0.9999    1.0000    0.2466    5.3386

              0       1
Pr(y|x)  0.7472  0.2528

           xgpa     tuce      psi
    x=  .311719  21.9375    .4375
sd(x)=  .046671  3.90151  .504016

. replace xgpa = gpa
. quietly logit grade xgpa tuce psi, nolog
. prchange xgpa

logit: Changes in Predicted Probabilities for grade

      min->max      0->1     -+1/2    -+sd/2  MargEfct
xgpa    0.7872    0.0008    0.5055    0.2466    0.5339

              0       1
Pr(y|x)  0.7472  0.2528

           xgpa     tuce      psi
    x=  3.11719  21.9375    .4375
sd(x)=  .466713  3.90151  .504016

. replace xgpa = gpa*10
. quietly logit grade xgpa tuce psi, nolog
. prchange xgpa

logit: Changes in Predicted Probabilities for grade

      min->max      0->1     -+1/2    -+sd/2  MargEfct
xgpa    0.7872    0.0000    0.0534    0.2466    0.0534

              0       1
Pr(y|x)  0.7472  0.2528

           xgpa     tuce      psi
    x=  31.1719  21.9375    .4375
sd(x)=  4.66713  3.90151  .504016
```