Margins Part 2: Marginal Effects

[References: Long 1997, Long and Freese 2003 & 2006, Cameron & Trivedi's "Microeconomics Using Stata" 2009 first edition.]

Overview. As Cameron & Trivedi note (p. 333), "An ME [marginal effect], or partial effect, most often measures the effect on the conditional mean of y of a change in one of the regressors, say X_k. In the linear regression model, the ME equals the relevant slope coefficient, greatly simplifying analysis. For nonlinear models, this is no longer the case, leading to remarkably many different methods for calculating MEs."

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. With binary dependent variables, they offer some of the same advantages that the Linear Probability Model (LPM) does – they give you a single number that expresses the effect of a variable on P(Y=1). The problem with the LPM, of course, is that the numbers are wrong, because relationships are not linear. With logistic regression and other nonlinear models, the marginal effects are "right" but you have to understand what they mean and what their limitations are.

Example. First we will show Marginal Effects at the Means (MEMS) and then later we'll do Average Marginal Effects (AMEs). As with our earlier discussion of predictive margins, with MEMs variables whose values are not explicitly fixed (e.g. race = 1) are set equal to their means when computing the marginal effects.

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

Logistic regression                               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
       1.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
------------------------------------------------------------------------------
```

```
. margins, dydx(*) atmeans

Conditional marginal effects                      Number of obs   =         32
Model VCE    : OIM

Expression   : Pr(grade), predict()
dy/dx w.r.t. : gpa tuce 1.psi
at           : gpa             =    3.117188 (mean)
               tuce            =     21.9375 (mean)
               0.psi           =       .5625 (mean)
               1.psi           =       .4375 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gpa |   .5338589    .237038     2.25   0.024      .069273    .9984447
        tuce |   .0179755   .0262369     0.69   0.493    -.0334479    .0693989
       1.psi |   .4564984   .1810537     2.52   0.012     .1016397    .8113571
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level
```

For MEMs, you can get similar output with the old `mfx` command. Long & Freese's `prchange` command also gives marginal effects. We'll now explain and discuss the numbers in the dy/dx column.

Discrete Change for Categorical Variables. Categorical variables, such as psi, can only take on two values, 0 and 1. It wouldn't make much sense to compute how P(Y=1) would change if, say, psi changed from 0 to .6, because that cannot happen. The MEM for categorical variables therefore shows how P(Y=1) changes as the categorical variable changes from 0 to 1, holding all other variables at their means. That is, for a categorical variable X_k

$$\text{Marginal Effect } X_k = \Pr(Y = 1|X, X_k = 1) - \Pr(y=1|X, X_k = 0)$$

In the current case, the MEM for psi of .456 tells us that, for two hypothetical individuals with average values on gpa (3.12) and tuce (21.94), the predicted probability of success is .456 greater for the individual in psi than for one who is in a traditional classroom. To confirm, we can easily compute the predicted probabilities for those hypothetical individuals:

```
. margins psi, atmeans

Adjusted predictions                              Number of obs   =         32
Model VCE    : OIM

Expression   : Pr(grade), predict()
at           : gpa             =    3.117188 (mean)
               tuce            =     21.9375 (mean)
               0.psi           =       .5625 (mean)
               1.psi           =       .4375 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         psi |
          0  |   .1067571   .0800945     1.33   0.183    -.0502252    .2637393
          1  |   .5632555   .1632966     3.45   0.001     .2432001    .8833109
------------------------------------------------------------------------------

. display .5632555 - .1067571
.4564984
```

For categorical variables with more than two possible values, e.g. religion, the marginal effects show you the difference in the predicted probabilities for cases in one category relative to the reference category. So, for example, if relig was coded 1 = Catholic, 2 = Protestant, 3 = Jewish, 4 = other, the marginal effect for Protestant would show you how much more (or less) likely Protestants were to succeed than were Catholics, the marginal effect for Jewish would show you how much more (or less) likely Jews were to succeed than were Catholics, etc.

Keep in mind that these are the marginal effects when all other variables equal their means (hence the term MEMs); the marginal effects will differ at other values of the Xs. For example, when gpa equals its minimum observed value and tuce equals its mean,

```
. margins , dydx(psi) at((min) gpa ) atmeans

Conditional marginal effects                      Number of obs   =         32
Model VCE    : OIM

Expression   : Pr(grade), predict()
dy/dx w.r.t. : 1.psi
at           : gpa             =        2.06 (min)
               tuce            =     21.9375 (mean)
               0.psi           =       .5625 (mean)
               1.psi           =       .4375 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       1.psi |   .0550463   .0761785     0.72   0.470    -.0942608    .2043534
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

The predicted benefit from being in psi is far smaller at this lower value of gpa, only 5.5 percent (compared to 45.6 percent when GPA equals its mean). That is, a C student would only see their chances of getting an A increase by 5.5 percent by being in psi, whereas a student with a low B+ average would see a 45.6 percent increase in their prospects by being in psi. This is consistent with what we have seen before: being in psi modestly improves the chances a C student has for getting an A, but psi has a much more beneficial effect for students in the B and above range.

Instantaneous rates of change for continuous variables. What does the MEM for gpa of .534 mean? It would be nice if we could say that a one unit increase in gpa will produce a .534 increase in the probability of success for an otherwise "average" individual. Sometimes statements like that will be (almost) true, but other times they won't. For example, if an "average" individual (average meaning gpa = 3.12, tuce = 21.94, psi = .4375) saw a one point increase in their gpa, here is how their predicted probability of success would change:

```
. margins, at(gpa = (3.117188 4.117188)) atmeans

Adjusted predictions                              Number of obs   =         32
Model VCE    : OIM

Expression   : Pr(grade), predict()

1._at        : gpa             =    3.117188
               tuce            =     21.9375 (mean)
               0.psi           =       .5625 (mean)
               1.psi           =       .4375 (mean)

2._at        : gpa             =    4.117188
               tuce            =     21.9375 (mean)
               0.psi           =       .5625 (mean)
               1.psi           =       .4375 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .2528205   .1052961     2.40   0.016      .046444     .459197
          2  |   .8510027   .1530519     5.56   0.000     .5510265    1.150979
------------------------------------------------------------------------------

. display .8510027 - .2528205
.5981822
```

Note that (a) the predicted increase of .598 is actually more than the MEM for gpa of .534, and (b) in reality, gpa couldn't go up 1 point for a person with an average gpa of 3.117.

MEMs for continuous variables measure the *instantaneous rate of change*, which may or may not be close to the effect on P(Y=1) of a one unit increase in X_k. Appendix 1 explains the concept in detail. What the MEM more or less tells you is that, if, say, X_k increased by some very small amount (e.g. .001), then P(Y=1) would increase by about .001*.534 = .000534, e.g.

```
. margins, at(gpa = (3.117188 3.118188)) atmeans noatlegend

Adjusted predictions                              Number of obs   =         32
Model VCE    : OIM

Expression   : Pr(grade), predict()

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .2528205   .1052961     2.40   0.016      .046444     .459197
          2  |   .2533547   .1053672     2.40   0.016     .0468388    .4598706
------------------------------------------------------------------------------

. display .2533547 - .2528205
.0005342
```

Put another way, for a continuous variable X_k,

$$\text{Marginal Effect of } X_k = \text{limit } [\Pr(Y = 1|X, X_k+\Delta) - \Pr(y=1|X, X_k)] / \Delta \ ]$$
$$\text{as } \Delta \text{ gets closer and closer to } 0$$

There is no guarantee that a bigger increase in X_k, e.g. 1, would produce an increase of 1*.534=.534. This is because the relationship between X_k and P(Y = 1) is nonlinear. When X_k is measured in small units, e.g. income in dollars, the effect of a 1 unit increase in X_k may match up well with the MEM for X_k. But, when X_k is measured in larger units (e.g. income in millions of dollars) the MEM may or may not provide a very good approximation of the effect of a one unit increase in X_k.

Discrete changes in all variables. So far, we have seen that (1) the values for marginal effects are dependent on the values of the X variables, and (2) the marginal effect for a continuous variable X_k may or may not be a good approximation of the effect that a 1 one unit increase in X_k will have on P(Y = 1). Put another way, the presentation of a single marginal effect for each X variable may or may not be helpful and informative in assessing the effect of changes in the X variables on P(Y=1). For these reasons, Long and others recommend examining predicted margins or marginal effects across a range of discrete values for one or more X variables (continuous or discrete). For example, we have already seen that the marginal effect of psi differs greatly between the lowest and mean values of gpa. Let us therefore examine the effect of changes in GPA on the marginal effect of psi in more detail (NOTE: I've edited the output to get rid of the redundant information on the means of tuce and psi):

```
. margins, dydx(psi) at(gpa = (1(.5)6)) atmeans vsquish

Conditional marginal effects                      Number of obs   =         32
Model VCE    : OIM

Expression   : Pr(grade), predict()
dy/dx w.r.t. : 1.psi
1._at        : gpa             =           1
               tuce            =     21.9375 (mean)
               0.psi           =       .5625 (mean)
               1.psi           =       .4375 (mean)
2._at        : gpa             =         1.5
3._at        : gpa             =           2
4._at        : gpa             =         2.5
5._at        : gpa             =           3
6._at        : gpa             =         3.5
7._at        : gpa             =           4
8._at        : gpa             =         4.5
9._at        : gpa             =           5
10._at       : gpa             =         5.5
11._at       : gpa             =           6
```

```
------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
1.psi        |
         _at |
          1  |   .0029386   .0079854     0.37   0.713    -.0127125    .0185897
          2  |   .0119416   .0250176     0.48   0.633     -.037092    .0609753
          3  |   .0469512    .068586     0.68   0.494    -.0874749    .1813772
          4  |   .1634819   .1368882     1.19   0.232    -.1048141    .4317779
          5  |   .4017712   .1738987     2.31   0.021      .060936    .7426065
          6  |   .5311925   .1902517     2.79   0.005     .1583061    .9040789
          7  |   .3482645   .2285081     1.52   0.127    -.0996032    .7961322
          8  |   .1285172   .1765621     0.73   0.467    -.2175381    .4745726
          9  |   .0355131   .0744611     0.48   0.633     -.110428    .1814543
         10  |   .0089342   .0245981     0.36   0.716    -.0392772    .0571455
         11  |   .0021924   .0074124     0.30   0.767    -.0123356    .0167205
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

The results show us that, for somebody who enters the course with a D gpa (_at = 1), being in psi rather than the regular class only improves the predicted chances of getting an A by 0.3 percent. Being in psi helps but it can't work miracles. A C student (_at = 3) does a little better, with psi improving their chances by 4.7 percent. Things really pick up for the B students (_at = 5) who see a 40.2 percent jump. Even straight A students (_at = 7) see a nice gain of 34.8 percent (which reflects the fact that even really smart students have a tough time in this class and hence can benefit from being in psi). If, however, there were somehow students out there with even higher gpas, the gains would be smaller; indeed, the genius with a 6.0 gpa would only gain 0.2 percent by being in psi.

Appendix 2 shows how Long & Freese's `prchange` command provides an alternative means for assessing the effects of discreet changes in variables.

Average Marginal Effects (AMEs). With MEMs, we use the mean values for any Xs whose values have not been fixed in the `margins` command. The use of means when computing marginal effects is criticized for the same reasons that means are criticized when computing predictive margins: (a) no real person may actually have mean values on all the Xs, (b) no real person has a value like .4375 on a categorical variable like psi (c) effects are only calculated at one set of values, the means. With the Average Marginal Effect, a marginal effect is computed for each case, and then all the computed effects are averaged. AMEs are the default for the `margins` command:

```
. quietly logit  grade gpa tuce i.psi
. margins, dydx(*)

Average marginal effects                          Number of obs   =         32
Model VCE    : OIM

Expression   : Pr(grade), predict()
dy/dx w.r.t. : gpa tuce 1.psi

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gpa |   .3625808   .1094411     3.31   0.001     .1480802    .5770815
        tuce |   .0122084   .0177942     0.69   0.493    -.0226675    .0470843
       1.psi |   .3575152   .1420034     2.52   0.012     .0791936    .6358367
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

Note that the AMES are somewhat different than the MEMs, suggesting that the marginal effect at the mean may not be a good reflection of the marginal effect at values other than the mean.

We still will find it helpful to see how the effect of psi differs across values of gpa:

```
. margins, dydx(psi) at(gpa = (1(.5)6)) vsquish

Average marginal effects                          Number of obs   =         32
Model VCE    : OIM

Expression   : Pr(grade), predict()
dy/dx w.r.t. : 1.psi
1._at        : gpa             =           1
2._at        : gpa             =         1.5
3._at        : gpa             =           2
4._at        : gpa             =         2.5
5._at        : gpa             =           3
6._at        : gpa             =         3.5
7._at        : gpa             =           4
8._at        : gpa             =         4.5
9._at        : gpa             =           5
10._at       : gpa             =         5.5
11._at       : gpa             =           6

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
1.psi        |
         _at |
          1  |   .0031263   .0086123     0.36   0.717    -.0137536    .0200062
          2  |   .0126789   .0269714     0.47   0.638    -.0401841    .0655418
          3  |   .0494715   .0731368     0.68   0.499    -.0938739     .192817
          4  |   .1683663    .139619     1.21   0.228    -.1052818    .4420145
          5  |   .3989144   .1696889     2.35   0.019     .0663302    .7314986
          6  |   .5189561   .1878684     2.76   0.006     .1507408    .8871714
          7  |   .3471089   .2180312     1.59   0.111    -.0802244    .7744423
          8  |   .1333261   .1828972     0.73   0.466    -.2251459    .4917981
          9  |   .0377277    .080625     0.47   0.640    -.1202943    .1957498
         10  |   .0095653   .0268938     0.36   0.722    -.0431455    .0622762
         11  |   .0023521   .0080983     0.29   0.771    -.0135202    .0182244
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

Conclusion. Marginal effects can be an informative means for summarizing how change in a response is related to change in a covariate (Stata 11 Reference Manual, p. 975). For categorical variables, the effects of discrete changes are computed, i.e., the marginal effects for categorical variables show how P(Y = 1) is predicted to change as X_k changes from 0 to 1 holding all other Xs equal. This can be quite useful, informative, and easy to understand.

For continuous independent variables, the marginal effect measures the instantaneous rate of change. If the instantaneous rate of change is similar to the change in P(Y=1) as X_k increases by one, this too can be quite useful and intuitive. However, there is no guarantee that this will be the case; it will depend, in part, on how X_k is scaled.

The effect of a change in X_k on P(Y=1) depends on the values of all of the X variables. It is therefore often useful to compute marginal effects at a range of values, e.g. see how the marginal effect of psi differs depending on someone's gpa.

Marginal Effects at the Means (MEMs) are computed by setting the values of X variables at their means, and then seeing how a change in one of the X_k variables changes P(Y = 1). With Average Marginal Effects (AMEs) a marginal effect is computed for each case, and the effects are then averaged. Many prefer AMEs because they think they provide a better representation of how changes in X_k affect P(Y = 1).

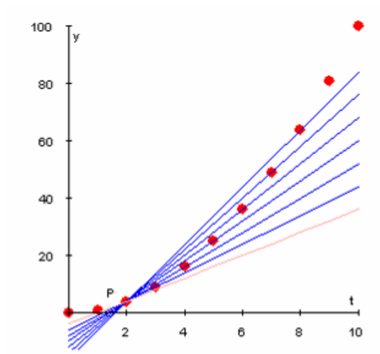
Appendix 1: Technical Discussion of Marginal Effects

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. But what is the slope of a curve??? 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. Again, 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 called 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* (MEM) is popular (i.e. compute the marginal effects when all x's are at their mean) but many think that *Average Marginal Effects* (AMEs) are superior.

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

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

Returning to our earlier example,

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

Logistic regression                               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
------------------------------------------------------------------------------

. adjust  gpa tuce psi, pr

--------------------------------------------------------------------------------------
     Dependent variable: grade     Equation: grade     Command: logit
 Covariates set to mean: gpa = 3.1171875, tuce = 21.9375, psi = .4375
--------------------------------------------------------------------------------------

----------------------
      All |         pr
----------+-----------
          |     .25282
----------------------
     Key:  pr  =  Probability

. 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
```

Looking specifically at GPA – 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 at the mean for GPA is therefore

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

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

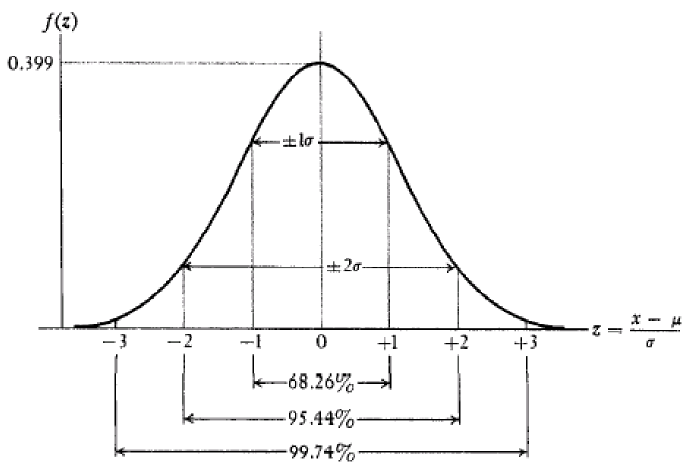
δ	GPA+δ	F(X,GPA+δ)	Change in F(X,GPA)	Change in GPA	Slope
10	13.11719	1	0.747179749	10	0.0747179749
1	4.11719	0.851002558	0.598182307	1	0.5981823074
0.1	3.21719	0.309808293	0.056988042	0.1	0.5698804165
0.05	3.16719	0.280431679	0.027611428	0.05	0.5522285640
0.01	3.12719	0.258196038	0.005375787	0.01	0.5375787403
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

(I think the `margins` command may actually do something similar to the above – it plugs in smaller and smaller values for delta until the change in slope is trivial.)

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

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

Using the `margins` command for MEMs & AMEs,

```
. quietly probit  grade gpa tuce i.psi, nolog
. margins, dydx(*) atmeans

Conditional marginal effects                      Number of obs   =         32
Model VCE    : OIM

Expression   : Pr(grade), predict()
dy/dx w.r.t. : gpa tuce 1.psi
at           : gpa             =    3.117188 (mean)
               tuce            =     21.9375 (mean)
               0.psi           =       .5625 (mean)
               1.psi           =       .4375 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gpa |   .5333471   .2324641     2.29   0.022     .0777259    .9889683
        tuce |   .0169697   .0271198     0.63   0.531    -.0361841    .0701235
       1.psi |    .464426   .1702807     2.73   0.006     .1306819    .7981701
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.

. margins, dydx(*)

Average marginal effects                          Number of obs   =         32
Model VCE    : OIM

Expression   : Pr(grade), predict()
dy/dx w.r.t. : gpa tuce 1.psi

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gpa |   .3607863   .1133816     3.18   0.001     .1385625    .5830102
        tuce |   .0114793   .0184095     0.62   0.533    -.0246027    .0475612
       1.psi |   .3737518   .1399913     2.67   0.008      .099374    .6481297
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

As a sidelight, note that the marginal effects (both MEMs and AMEs) for probit are very similar to the marginal effects for logit. This is usually the case.

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
------------------------------------------------------------------------------

. margins, dydx(*) atmeans

Conditional marginal effects                      Number of obs   =        500
Model VCE    : OLS

Expression   : Linear prediction, predict()
dy/dx w.r.t. : educ jobexp black
at           : educ            =       13.16 (mean)
               jobexp          =       13.52 (mean)
               black           =          .2 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   1.840407   .0467507    39.37   0.000     1.748777    1.932036
      jobexp |   .6514259   .0350604    18.58   0.000     .5827087    .7201431
       black |   -2.55136   .4736266    -5.39   0.000    -3.479651   -1.623069
------------------------------------------------------------------------------

. margins, dydx(*)

Average marginal effects                          Number of obs   =        500
Model VCE    : OLS

Expression   : Linear prediction, predict()
dy/dx w.r.t. : educ jobexp black

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   1.840407   .0467507    39.37   0.000     1.748777    1.932036
      jobexp |   .6514259   .0350604    18.58   0.000     .5827087    .7201431
       black |   -2.55136   .4736266    -5.39   0.000    -3.479651   -1.623069
------------------------------------------------------------------------------
```

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.

Appendix 2: The prchange command.

Long & Freese's `prchange` command conveniently provides several useful measures of discrete change. These aren't marginal effects, but they do offer many of the same advantages. All of these calculations could also be done using `margins`, but it would take some work. (Note, however, that `prchange` will set the values of X variables at their means, whereas `margins` can use asobserved.) 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, `prchange` 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 at their means). Example:

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

logit: Changes in 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

 Pr(y|x): probability of observing each y for specified x values
Avg|Chg|: average of absolute value of the change across categories
Min->Max: change in predicted probability as x changes from its minimum to
          its maximum
    0->1: change in predicted probability as x changes from 0 to 1
   -+1/2: change in predicted probability as x changes from 1/2 unit below
          base value to 1/2 unit above
  -+sd/2: change in predicted probability as x changes from 1/2 standard
          dev below base to 1/2 standard dev above
MargEfct: the partial derivative of the predicted probability/rate with
          respect to a given independent variable
```

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 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.

Similarly, when gpa equals its lowest observed value (2.06) and the other variables equal their means, the probability of success is only 1.68 percent. But, when gpa equals its largest observed value (4.00) and the other variables equal their means, the probability of success is 80.4 percent, a jump of 78.72 percent. If we wanted to do the min/max calculation with `margins`,

```
. margins, at((min) gpa) at((max) gpa) atmeans

Adjusted predictions                              Number of obs   =         32
Model VCE    : OIM

Expression   : Pr(grade), predict()

1._at        : gpa             =        2.06 (min)
               tuce            =     21.9375 (mean)
               0.psi           =       .5625 (mean)
               1.psi           =       .4375 (mean)

2._at        : gpa             =           4 (max)
               tuce            =     21.9375 (mean)
               0.psi           =       .5625 (mean)
               1.psi           =       .4375 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .0167682   .0264193     0.63   0.526    -.0350126    .0685491
          2  |    .803971   .1695703     4.74   0.000     .4716193    1.136323
------------------------------------------------------------------------------

. display .803971 - .0167682
.7872028
```