POLS 6481. Research Design and Quantitative Methods II
Lecture 18. Binary Dependent Variables – Probit and Logit
Readings: Wooldridge, *Introductory Econometrics 5e*, 17.1

Outline

0. Prelude: Wolf Spiders
1. Latent Variable Formulation
2. Probit
3. Logit
4. Comparing Probit and Logit
5. R Code for Estimating 'Generalized Linear Models'
6. R Code for Interpreting 'Generalized Linear Models'

POLS 6481. Research Design and Quantitative Methods II
Lecture 18. Binary Dependent Variables – Probit and Logit
Readings: Wooldridge, *Introductory Econometrics 5e*, 17.1

0. Prelude: *Wolf Spiders*

Data: wolfspiders.csv
R Script: Lecture 18 wolf spiders.R

Based on the article, "Sexual Cannibalism and Mate Choice Decisions in Wolf Spiders: Influence of Male Size and Secondary Sexual Characteristics"

N = 52 couples

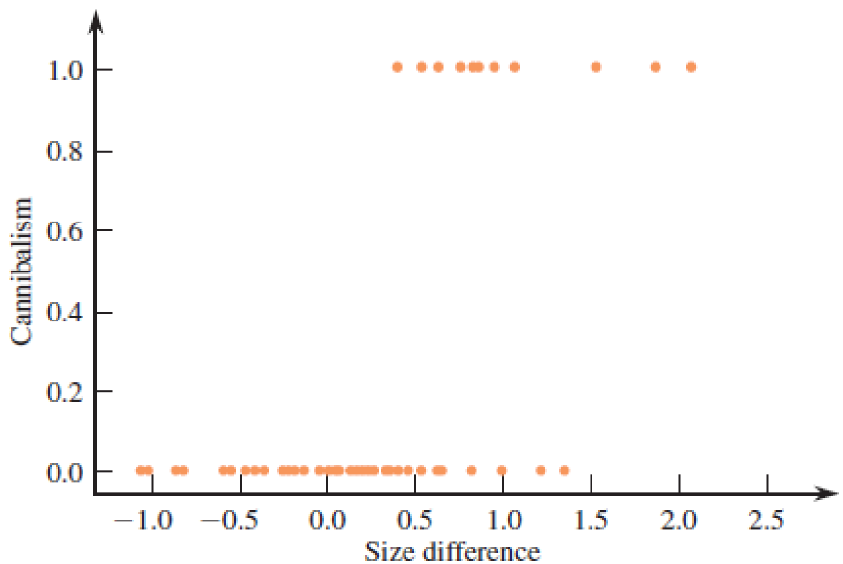
Dependent variable: whether female wolf spider kills and consumes her partner during courtship or mating (11 events coded 1, 41 events coded 0)

Independent variable: *size difference* = female body width minus male body width

POLS 6481. Research Design and Quantitative Methods II
Lecture 18. Binary Dependent Variables – Probit and Logit
Readings: Wooldridge, *Introductory Econometrics 5e*, 17.1

0. Prelude: *Wolf Spiders*

Size Difference (mm)	Cannibalism	Size Difference (mm)	Cannibalism
−1	0	0.4	0
−1	0	0.4	0
−0.8	0	0.4	0
−0.8	0	0.4	0
−0.6	0	0.4	1
−0.6	0	0.6	0
−0.4	0	0.6	0
−0.4	0	0.6	0
−0.4	0	0.6	0
−0.4	0	0.6	0
−0.2	0	0.6	1
−0.2	0	0.6	1
−0.2	0	0.8	0
−0.2	0	0.8	0
0.0	0	0.8	1
0.0	0	0.8	1
0.0	0	0.8	1
0.0	0	1.0	0
0.0	0	1.0	0
0.0	0	1.0	1
0.2	0	1.0	1
0.2	0	1.2	0
0.2	0	1.4	0
0.2	0	1.6	1
0.2	0	1.8	1
0.2	0	2.0	1



POLS 6481. Research Design and Quantitative Methods II
Lecture 18. Binary Dependent Variables – Probit and Logit
Readings: Wooldridge, *Introductory Econometrics 5e*, 17.1

0. Prelude: *Wolf Spiders*

Data: wolfspiders.csv
R Script: Lecture 18 wolf spiders.R

OLS – LPM equation:

$$\underset{(.05287)}{0.10679} + \underset{(.07109)}{0.33623}\cdot \textit{Size Difference}$$

Logit equation:

$$\underset{(.82878)}{-3.08904} + \underset{(1.00407)}{3.06928}\cdot \textit{Size Difference}$$

POLS 6481. Research Design and Quantitative Methods II
Lecture 18. Binary Dependent Variables – Probit and Logit
Readings: Wooldridge, *Introductory Econometrics 5e*, 17.1

1. The Latent Variable Formulation

Consider the decision of whether to smoke; assume an individual *i*'s decision depends on his/her own (unobservable) utility index I_i^*, which depends on explanatory variables like education, the price of cigarettes, income, …

Let $I_i^* = \beta\mathbf{x} + u_i$

Suppose that a person smokes ($Y_i = 1$) if $I_i^* > 0$; and a person does not smoke ($Y_i = 0$) if $I_i^* \leq 0$

Then the probability that a person smokes $Pr(Y_i = 1) = Pr(I_i^* > 0)$

$$= Pr(\beta\mathbf{x}_i + u_i \geq 0)$$
$$= Pr(\, u_i \geq -\beta\mathbf{x}_i)$$

If this probability distribution is symmetric, then we can re-write the last equality as:

$$Pr(Y_i = 1) = Pr(u_i \leq \beta\mathbf{x}_i)$$

POLS 6481. Research Design and Quantitative Methods II
Lecture 18. Binary Dependent Variables – Probit and Logit
Readings: Wooldridge, *Introductory Econometrics 5e*, 17.1

1. The Latent Variable Formulation

Wooldridge 17.1

$$y^* = \boldsymbol{x\beta} + e$$

$$y = 1\,[y^* > 0]$$

$$\Rightarrow P(y=1|\mathbf{x}) = P(y^* > 0|\mathbf{x})$$

$$= P(e > -\boldsymbol{x\beta}) = 1 - \boxed{G(-\boldsymbol{x\beta})} = \boxed{G(\boldsymbol{x\beta})}$$

If the latent variable y^* exceeds zero, then y takes on the value 1.
(If y^* is less than or equal to zero, then y takes on the value 0.)
Thus, y^* can be interpreted as the propensity to have $y = 1$.

Marginal effects

$$\frac{\partial P(y=1|\mathbf{x})}{\partial x_j} = \boxed{g(\mathbf{x}\boldsymbol{\beta})}\beta_j \text{ where } g(z) = \boxed{\partial G(z)}/\partial z > 0$$

For this semester, don't worry a lot about $G(\boldsymbol{x\beta})$ and $g(\boldsymbol{x\beta})$

Key question: How does the probability of y equaling 1 change as values of x_j change?

Answer: Unfortunately, the effects are not constant across values of x_j.

Even worse answer: the effects are not constant across values of the other x's!

POLS 6481. Research Design and Quantitative Methods II
Lecture 18. Binary Dependent Variables – Probit and Logit
Readings: Wooldridge, *Introductory Econometrics 5e*, 17.1

2. Probit

$$\boxed{G(z)} = \Phi(z) = \int_{-\infty}^{z} \phi(v)dv$$

Right now, you shouldn't care how the model is estimated; you should care about what to do with the coefficients once you have them. The simple answer is:

1. Check coefficients for signs (positive ↔ increases probability $y = 1$)
2. Check coefficients for statistical significance (z replaces t; the critical value = 1.96)
3. Assign values to all $x_1, \ldots, x_k$
4. Using the estimated coefficients, compute the linear combination: $\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$
5. Plug your answer into the standard normal distribution* for the predicted probability
6. Choose one variable, x_j, and change its value from the baseline to measure its impact
7. Repeat steps 3–5 until you've done enough "comparative statics"
8. Make a table of predicted probabilities and (conditional) marginal effects

* The standard normal distribution tells you $\int_{-\infty}^{x\beta} \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{t}{2}\right)^2} dt$ where $t = x\beta$

A shortcut to step 4 is two-sided handout with tables translating $\hat{z}$ into $\widehat{pr}(y = 1)$

Readings: Wooldridge, *Introductory Econometrics 5e*, 17.1

3. Logit

$$\boxed{G(z)} = \Lambda(z) = \exp(z)/[1 + \exp(z)]$$

As before, once you have estimated coefficients:

1. Check coefficients for signs (positive ↔ increases probability $y = 1$)
2. Check coefficients for statistical significance ($z > 1.96$?)
3. Assign values to all $x_1, \ldots, x_k$
4. Using the estimated coefficients, compute the linear combination:
5. Plug your answer into this equation: $\frac{1}{1+e^{-z}}$ where $z = x\beta$
6. Choose one variable, x_j, and change its value from the baseline to measure its impact
7. Repeat steps 3–5 until you've done enough "comparative statics"
8. Make a table of predicted probabilities and (conditional) marginal effects

A shortcut to step 4 is two-sided handout with tables translating $\hat{z}$ into $\widehat{pr}(y = 1)$

3. Logit

$$\boxed{G(z)} = \Lambda(z) = \exp(z)/[1 + \exp(z)]$$

A lot of people find logit easier to work with logit than probit; the basic ideas are:

$P_i = \dfrac{1}{1+e^{-Z_i}}$ where $Z_i = \beta\mathbf{x}_i + u_i$

$1-P_i = \ldots = \dfrac{1}{1+e^{Z_i}}$ where $Z_i = \beta\mathbf{x}_i + u_i$

$\dfrac{P_i}{1-P_i} = \dfrac{1+e^{Z_i}}{1+e^{-Z_i}} = \ldots = e^{Z_i}$ (as P_i goes from 0 to 1, the “odds ratio” goes from 0 to $+\infty$)

$L_i = ln\,\dfrac{P_i}{1-P_i} = ln\; e^{Z_i} = Z_i = \beta\mathbf{x}_i + u_i$ (as P_i goes from 0 to 1, the “logit” goes from $-\infty$ to $+\infty$)

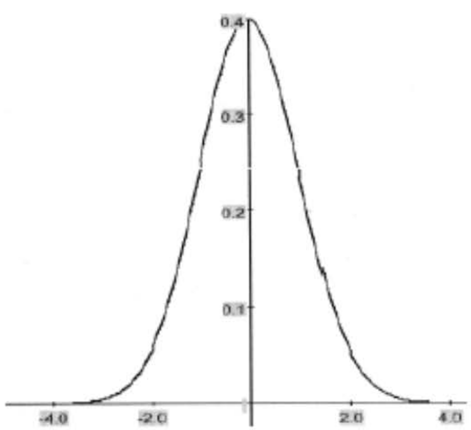
4. Comparing Probit and Logit

Logit coefficients are usually larger than probit coefficients, but their standard errors are too, so significance tests rarely change.

Logit is easier to derive (start with the odds ratio, then the log odds ratio, then…)

Logit has slightly thicker tails than probit.

PDF of Probit:



PDF of Logit:

POLS 6481. Research Design and Quantitative Methods II
Lecture 18. Binary Dependent Variables – Probit and Logit
Readings: Wooldridge, *Introductory Econometrics 5e*, 17.1

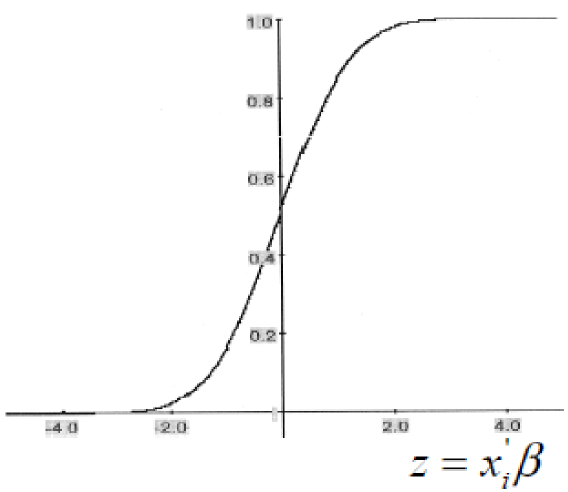
4. Comparing Probit and Logit

Logit coefficients are usually larger than probit coefficients, but their standard errors are too, so significance tests rarely change.

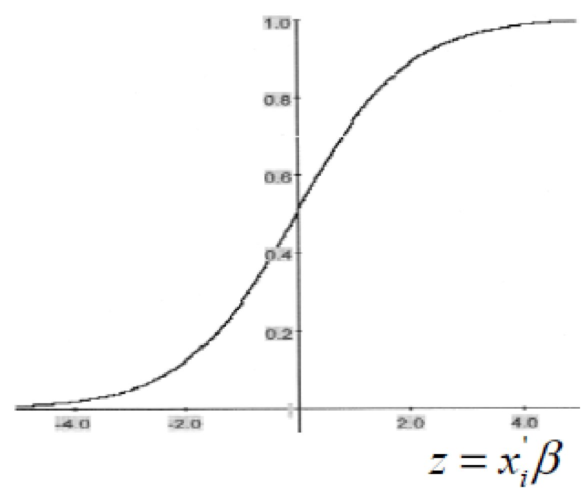
Logit is easier to derive (start with the odds ratio, then the log odds ratio, then…)

Logit has slightly thicker tails than probit.

CDF of Probit:



CDF of Logit:

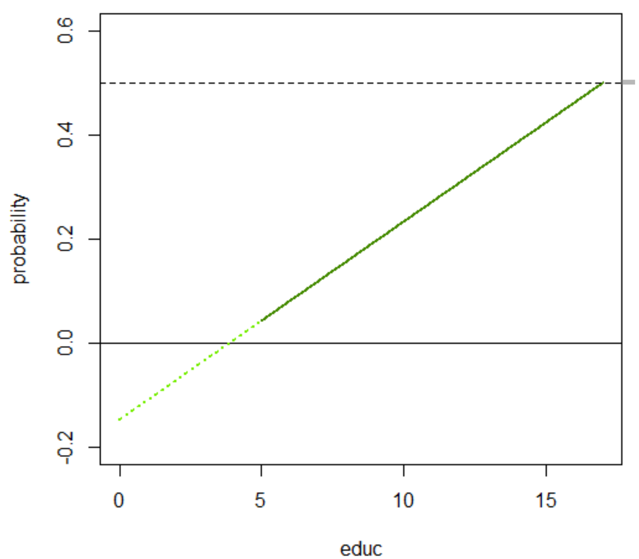


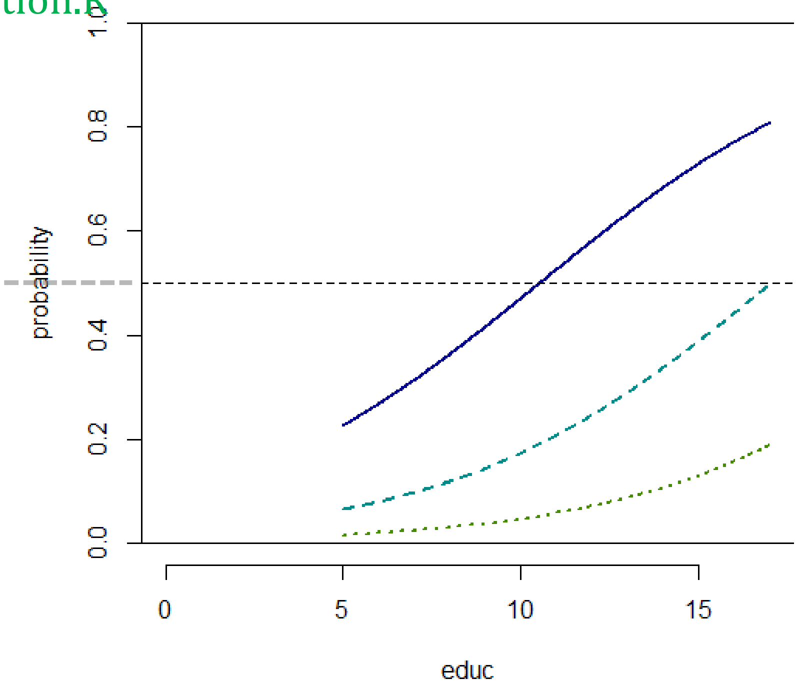
POLS 6481. Research Design and Quantitative Methods II
Lecture 18. Binary Dependent Variables – Probit and Logit
Readings: Wooldridge, *Introductory Econometrics 5e*, 17.1

4. Comparing Logit and Probit (and LPM)

Example of Labor force participation for married women

Data: MROZ.DTA
R Script: Lecture 17+18 labor force particpation.R





4. Comparing Logit and Probit (and LPM)

Example of Labor force participation for married women

TABLE 17.1 LPM, Logit, and Probit Estimates of Labor Force Participation

Dependent Variable: *inlf*			
Independent Variables	**LPM (OLS)**	**Logit (MLE)**	**Probit (MLE)**
nwifeinc	−.0034 (.0015)	−.021 (.008)	−.012 (.005)
educ	.038 (.007)	.221 (.043)	.131 (.025)
exper	.039 (.006)	.206 (.032)	.123 (.019)
$exper^2$	−.00060 (.00018)	−.0032 (.0010)	−.0019 (.0006)
age	−.016 (.002)	−.088 (.015)	−.053 (.008)
kidslt6	−.262 (.032)	−1.443 (.204)	−.868 (.119)
kidsge6	.013 (.013)	.060 (.075)	.036 (.043)
constant	.586 (.151)	.425 (.860)	.270 (.509)
Percentage correctly predicted	73.4	73.6	73.4
Log-likelihood value	—	−401.77	−401.30
Pseudo *R*-squared	.264	.220	.221

Coefficients are not comparable across models. Logit estimated coefficients are 1.6 times as large as Probit estimated coefficients, because:

$$g_{Logit}(0)/g_{Probit}(0) \approx 1/1.6$$

The biggest difference between the LPM and Logit/Probit is that partial effects are not constant in Logit/Probit:

$$\widehat{P}(working|\bar{\boldsymbol{x}}, kidslt6 = 0) = .707$$

$$\widehat{P}(working|\bar{\boldsymbol{x}}, kidslt6 = 1) = .373$$

$$\widehat{P}(working|\bar{\boldsymbol{x}}, kidslt6 = 2) = .117$$

(Larger decrease in probability for the first child)

5. R Code for Estimating 'Generalized Linear Models'

LPM: `lpm <- glm(y ~ x1 + x2, family=gaussian); summary(lpm)`

Logit: `logit <- glm(y ~ x1 + x2, binomial(link = "logit")); summary(logit)`

Probit: `probit <- glm(y ~ x1 + x2, binomial(link = "probit")); summary(probit)`

6. R Code for Interpreting 'Generalized Linear Models'

Predicted Values from GLM's

LPM: `lpm$coef [1] + x1*lpm$coef[2] + x2*lpm$coef[3]`

Logit: `(1+exp(-(logit$coef[1] + x1*logit$coef[2] + x2*logit$coef[3])))^-1`

Probit: `pnorm(probit$coef[1] + x1*probit$coef[2] + x2*probit$coef[3])`

You would need to insert or define specific values of x1 and x2

Marginal Effects from GLM's

The *erer* package has a function *maBina* that allows you to compute marginal effects with each regressor set at its mean value. For example:

```
maBina(logit,  x.mean = T, rev.dum=T, digits=3, subset.name=NA)
maBina(probit,  x.mean = T, rev.dum=T, digits=3, subset.name=NA)
```

To use maBina without the erer package, you must (1) run two scripts, maBina.r & listn.r.
To use maBina re-run glm() *with* x = TRUE, *<u>added</u> inside parentheses*

POLS 6481. Research Design and Quantitative Methods II
Lecture 18. Binary Dependent Variables – Probit and Logit
Readings: Wooldridge, *Introductory Econometrics 5e*, 17.1

6. R Code for Interpreting 'Generalized Linear Models'

Predicted Values from GLM's

LPM: `lpm$coef [1] + x1*lpm$coef[2] + x2*lpm$coef[3]`

Logit: `(1+exp(-(logit$coef[1] + x1*logit$coef[2] + x2*logit$coef[3])))^-1`

Probit: `pnorm(probit$coef[1] + x1*probit$coef[2] + x2*probit$coef[3])`

You would need to insert specific values of x1 and x2

Marginal Effects from GLM's

Or you can plot the marginal effects with confidence intervals…
A few years ago I wrote an R script, but the "mfxboot" program works really slowly…

6. R Code for Interpreting 'Generalized Linear Models'

Example: Effect of 'Personalized System of Instruction'

Data: logist.dta
R Script: Lecture 18 plotting without erer.R

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

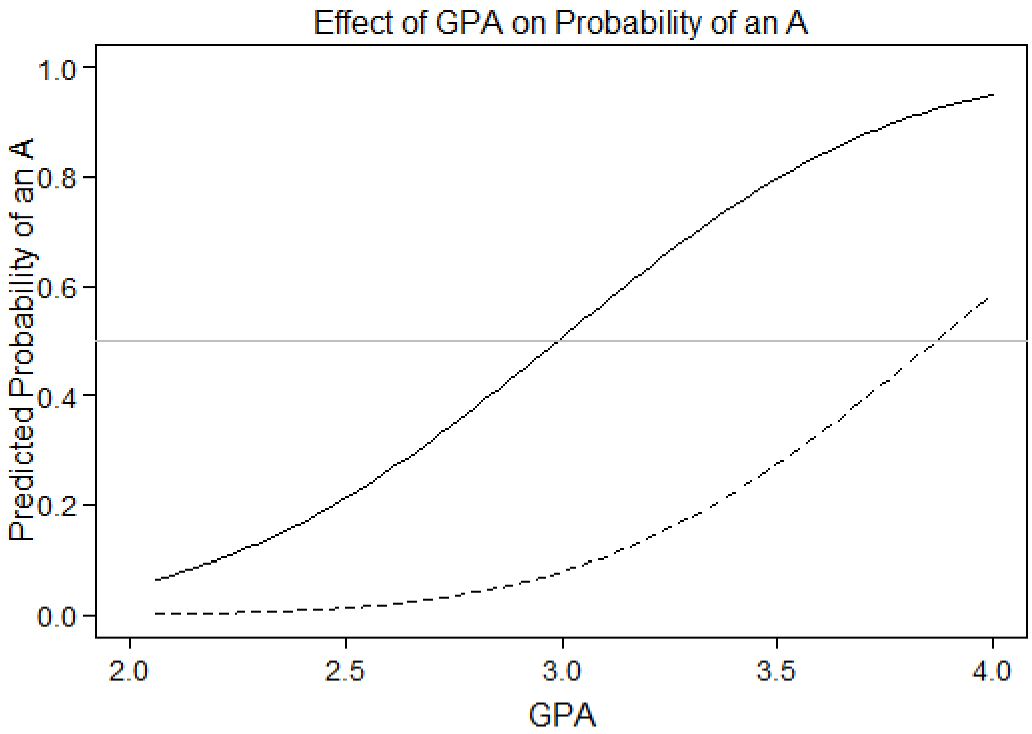
grade	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
psi	1.426332	.595037	2.40	0.017	.2600814	2.592583
tuce	.0517289	.0838901	0.62	0.537	-.1126927	.2161506
gpa	1.62581	.6938818	2.34	0.019	.2658269	2.985794
_cons	-7.45232	2.542467	-2.93	0.003	-12.43546	-2.469177

```
probit <- glm(grade ~ psi + tuce + gpa, x = TRUE, data, family=binomial(link = "probit"),
na.action=na.exclude); summary(probit)
```

POLS 6481. Research Design and Quantitative Methods II
Lecture 18. Binary Dependent Variables – Probit and Logit
Readings: Wooldridge, *Introductory Econometrics 5e*, 17.1

6. R Code for Interpreting 'Generalized Linear Models'

Example: Effect of 'Personalized System of Instruction'



POLS 6481. Research Design and Quantitative Methods II
Lecture 18. Binary Dependent Variables – Probit and Logit
Readings: Wooldridge, *Introductory Econometrics 5e*, 17.1

6. R Code for Interpreting ‘Generalized Linear Models’

Example: Effect of ‘Personalized System of Instruction’

```
Marginal effects on Prob(grade==1) after probit

------------------------------------------------------------------------------
       grade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gpa |   .3637883   .1129461     3.22   0.001     .1424181    .5851586
        tuce |    .011476   .0184085     0.62   0.533    -.024604     .047556
         psi |   .3737518   .1399912     2.67   0.008     .0993741    .6481295
------------------------------------------------------------------------------
```

I don’t know where this table ⇧ came from! Maybe someone’s lecture notes (using Stata)? Anyways, here’s the marginal effects for using the *maBina* program in the *erer* package:

```
maBina(probit, x.mean = T, rev.dum = T)
maBina(probit, x.mean = T, rev.dum = T, subset.name = "psi", subset.value = 1)
maBina(probit, x.mean = T, rev.dum = T, subset.name = "psi", subset.value = 0)
```