Binary Dependent Variables

In some cases the outcome of interest – rather than one of the right hand side variables - is discrete rather than continuous

Binary Dependent Variables

In some cases the outcome of interest – rather than one of the right hand side variables - is discrete rather than continuous

The simplest example of this is when the Y variable is binary – so that it can take only 1 or 2 possible values (eg Pass/Fail, Profit/Loss, Win/Lose)

Binary Dependent Variables

In some cases the outcome of interest – rather than one of the right hand side variables - is discrete rather than continuous

The simplest example of this is when the Y variable is binary – so that it can take only 1 or 2 possible values (eg Pass/Fail, Profit/Loss, Win/Lose)

Representating this is easy – just use a dummy variable on the left hand side of your model, so that

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \qquad (1)$$

Binary Dependent Variables

In some cases the outcome of interest – rather than one of the right hand side variables - is discrete rather than continuous

The simplest example of this is when the Y variable is binary – so that it can take only 1 or 2 possible values (eg Pass/Fail, Profit/Loss, Win/Lose)

Representating this is easy – just use a dummy variable on the left hand side of your model, so that

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \qquad (1)$$

Becomes

$$D_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \qquad (2)$$

Binary Dependent Variables

In some cases the outcome of interest – rather than one of the right hand side variables - is discrete rather than continuous

The simplest example of this is when the Y variable is binary – so that it can take only 1 or 2 possible values (eg Pass/Fail, Profit/Loss, Win/Lose)

Representating this is easy – just use a dummy variable on the left hand side of your model, so that

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \qquad (1)$$

Becomes

$$D_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \qquad (2)$$

Where D_i = 1 if the individual (or firm or country etc) has the characteristic (eg Win)

Binary Dependent Variables

In some cases the outcome of interest – rather than one of the right hand side variables - is discrete rather than continuous

The simplest example of this is when the Y variable is binary – so that it can take only 1 or 2 possible values (eg Pass/Fail, Profit/Loss, Win/Lose)

Representating this is easy – just use a dummy variable on the left hand side of your model, so that

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \qquad (1)$$

Becomes

$$D_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \qquad (2)$$

Where D_i = 1 if the individual (or firm or country etc) has the characteristic (eg Win)

= 0 if not (eg Lose)

$$D_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \qquad (2)$$

Equation (2) can be estimated by OLS.

$$D_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \qquad (2)$$

Equation (2) can be estimated by OLS.

When do this it is called a **linear probability model** and can interpret the coefficients in a similar way as with other OLS models

(linear because if fits a straight line and probability because it implicitly models the probability of an event occurring)

$$D_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \qquad (2)$$

Equation (2) can be estimated by OLS.

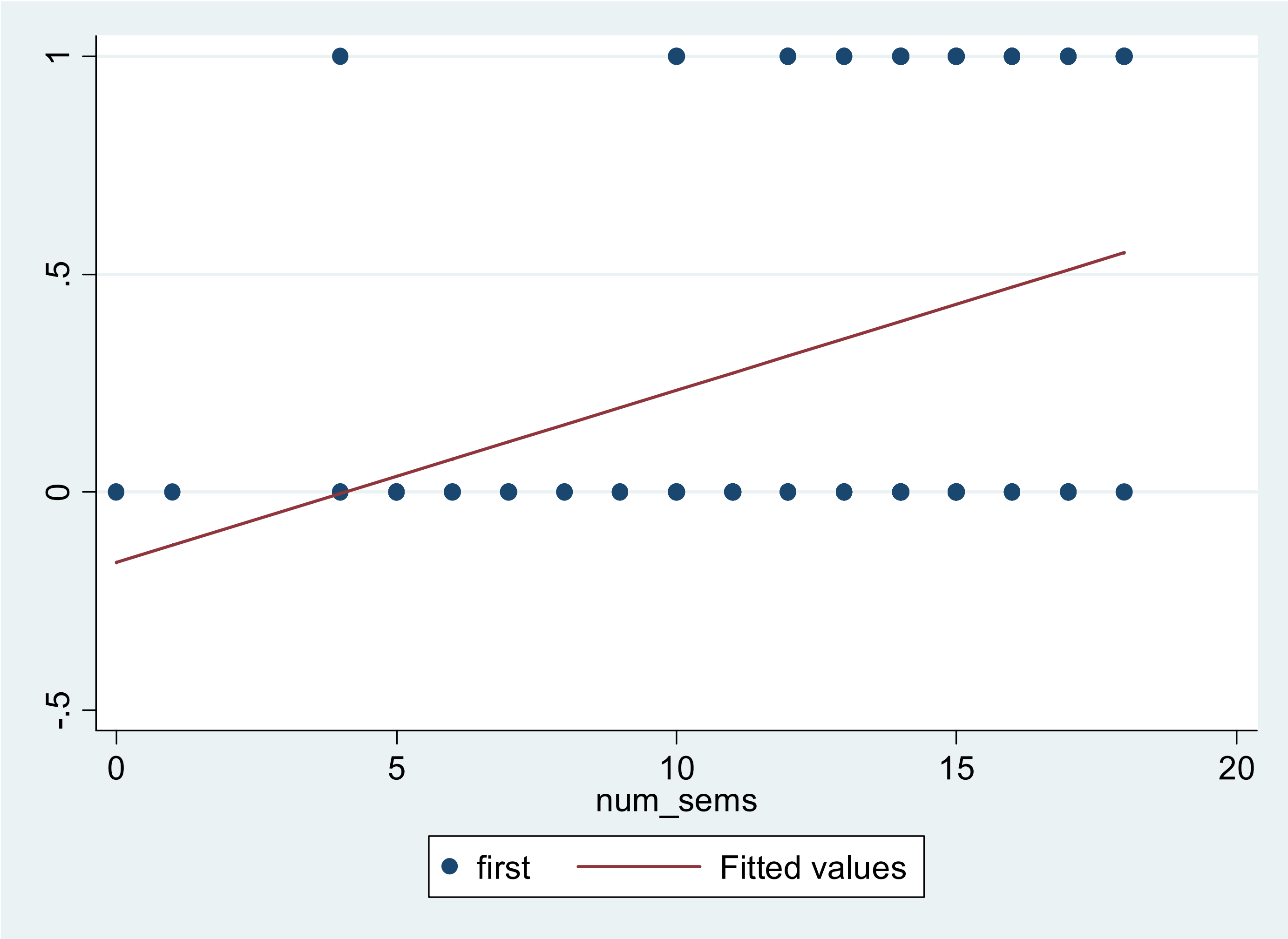
When do this it is called a **linear probability model** and can interpret the coefficients in a similar way as with other OLS models

$$D_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \qquad (2)$$

Equation (2) can be estimated by OLS.

When do this it is called a **linear probability model** and can interpret the coefficients in a similar way as with other OLS models

(linear because if fits a straight line and probability because it implicitly models the probability of an event occurring)



two (scatter first num_sems) (line phat num_sems)

The chance of getting a first can be seen to depend positively on the number of seminars in this linear probability model

$$D_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \qquad (2)$$

Equation (2) can be estimated by OLS.

When do this it is called a **linear probability model** and can interpret the coefficients in a similar way as with other OLS models

(linear because if fits a straight line and probability because it implicitly models the probability of an event occurring)

So for example the coefficient β_1

$$D_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \qquad (2)$$

Equation (2) can be estimated by OLS.

When do this it is called a **linear probability model** and can interpret the coefficients in a similar way as with other OLS models

(linear because if fits a straight line and probability because it implicitly models the probability of an event occurring)

So for example the coefficient $\beta_1 = dD_i/dX_1$

$$D_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \qquad (2)$$

Equation (2) can be estimated by OLS.

When do this it is called a **linear probability model** and can interpret the coefficients in a similar way as with other OLS models

(linear because if fits a straight line and probability because it implicitly models the probability of an event occurring)

So for example the coefficient $\beta_1 = dD_i/dX_1$

now gives the impact of a unit change in the value of X_1 on the chances of belonging to the category coded D=1 (eg of winning) - hence the name linear probability model

Example: Using the dataset marks.dta can work out the chances of getting a first depend on class attendance and gender using a linear probability model

```
gen first=mark>=70
/* first set up a dummy variable that will become the dependent variable */
tab first
      first |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |         79       66.95       66.95
          1 |         39       33.05      100.00
------------+-----------------------------------
      Total |        118      100.00
```

So 39 students got a first out of 118. If we summarise this variable then the mean value of this (or any) binary variable is the proportion of the sample with that characteristic

```
ci first
    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
       first |        118    .3305085    .0434881        .2443825    .4166345
```

So in this case 33% of the course got a first class mark (.33 ≡ 33%)
To see what determines this look at the OLS regression output

```
reg first num_sems female

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F(  2,   115) =   12.03
       Model |  4.51869886     2  2.25934943           Prob > F      =  0.0000
    Residual |  21.5914706   115  .187751919           R-squared     =  0.1731
-------------+------------------------------           Adj R-squared =  0.1587
       Total |  26.1101695   117  .223163842           Root MSE      =   .4333
------------------------------------------------------------------------------
       first |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    num_sems |   .0342389   .0096717     3.54   0.001     .0150811    .0533967
      female |   .2144069   .0837238     2.56   0.012     .0485662    .3802475
       _cons |  -.1796094   .1244661    -1.44   0.152    -.4261528     .066934
```

This says that the chances of getting a first rise by 3.4 percentage points for every extra class attended
and that, on average, women are 21 percentage points more likely to get a first, even after allowing for the number of classes attended.

However, can show that OLS estimates when the dependent variable is binary

However, can show that OLS estimates when the dependent variable is binary

1. will suffer from heteroskedasticity, so that the t-statistics are biased

However, can show that OLS estimates when the dependent variable is binary

1. will suffer from heteroskedasticity, so that the t-statistics are biased

2. as graph shows may not constrain the predicted values to lie between 0 and 1 (which need if going to predict behaviour accurately)

Using the example above

```
predict phat
(option xb assumed; fitted values)

. su phat

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        phat |       118    .3305085    .1965232  -.1796094   .6510978
```

Can see that there are some negative predictions which is odd for a variable that is binary. Note that not many of the predicted values are zero (and none are 1). This is not unusual since the model is actually predicting the probability of belonging to one of the two categories.

Because of this it is better to use a model that explicitly rather than implicitly models this probability and does not suffer from heteroskedasticity

Because of this it is better to use a model that explicitly rather than implicitly models this probability and does not suffer from heteroskedasticity

So that model

Probability (D_i=1)

Because of this it is better to use a model that explicitly rather than implicitly models this probability and does not suffer from heteroskedasticity

So that model

$$\text{Probability } (D_i=1)$$

as a function of the right hand side variables

Because of this it is better to use a model that explicitly rather than implicitly models this probability and does not suffer from heteroskedasticity

So that model

$$\text{Probability } (D_i=1)$$

as a function of the right hand side variables

rather than simply measure whether D=1 or 0

Because of this it is better to use a model that explicitly rather than implicitly models this probability and does not suffer from heteroskedasticity

So that model

$$\text{Probability } (D_i=1)$$

as a function of the right hand side variables

rather than simply measure whether D=1 or 0

$$\text{Probability } (D_i=1) = F(X_1, X_2 \ldots X_k)$$

Because of this it is better to use a model that explicitly rather than implicitly models this probability and does not suffer from heteroskedasticity

So that model

$$\text{Probability } (D_i=1)$$

as a function of the right hand side variables

rather than simply measure whether D=1 or 0

$$\text{Probability } (D_i=1) = F(X_1, X_2 \ldots X_k)$$

So need to come up with a specific functional form for this relationship

There are 2 alternatives commonly used

1. Logit Model

Assumes that the Probability model

$$\text{Probability } (D_i=1) = F(X_{1\ ,}\ X_2 \ldots\ X_k)$$

is given by

1. Logit Model

Assumes that the Probability model

$$\text{Probability } (D_i=1) = F(X_1, X_2 \ldots X_k)$$

is given by

$$\text{Prob } (D_i=1) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2)}}$$

1. Logit Model

Assumes that the Probability model

$$\text{Probability } (D_i=1) = F(X_1, X_2 \ldots X_k)$$

is given by

$$\text{Prob } (D_i=1) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2)}}$$

2. The Probit Model

Assumes that the Probability model is given by

1. Logit Model

Assumes that the Probability model

$$\text{Probability } (D_i=1) = F(X_1, X_2 \ldots X_k)$$

is given by

$$\text{Prob } (D_i=1) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2)}}$$

2. The Probit Model

Assumes that the Probability model is given by

$$\text{Prob } (D_i=1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\beta_0+\beta_1 X_1+\beta_2 X_2)^2}$$

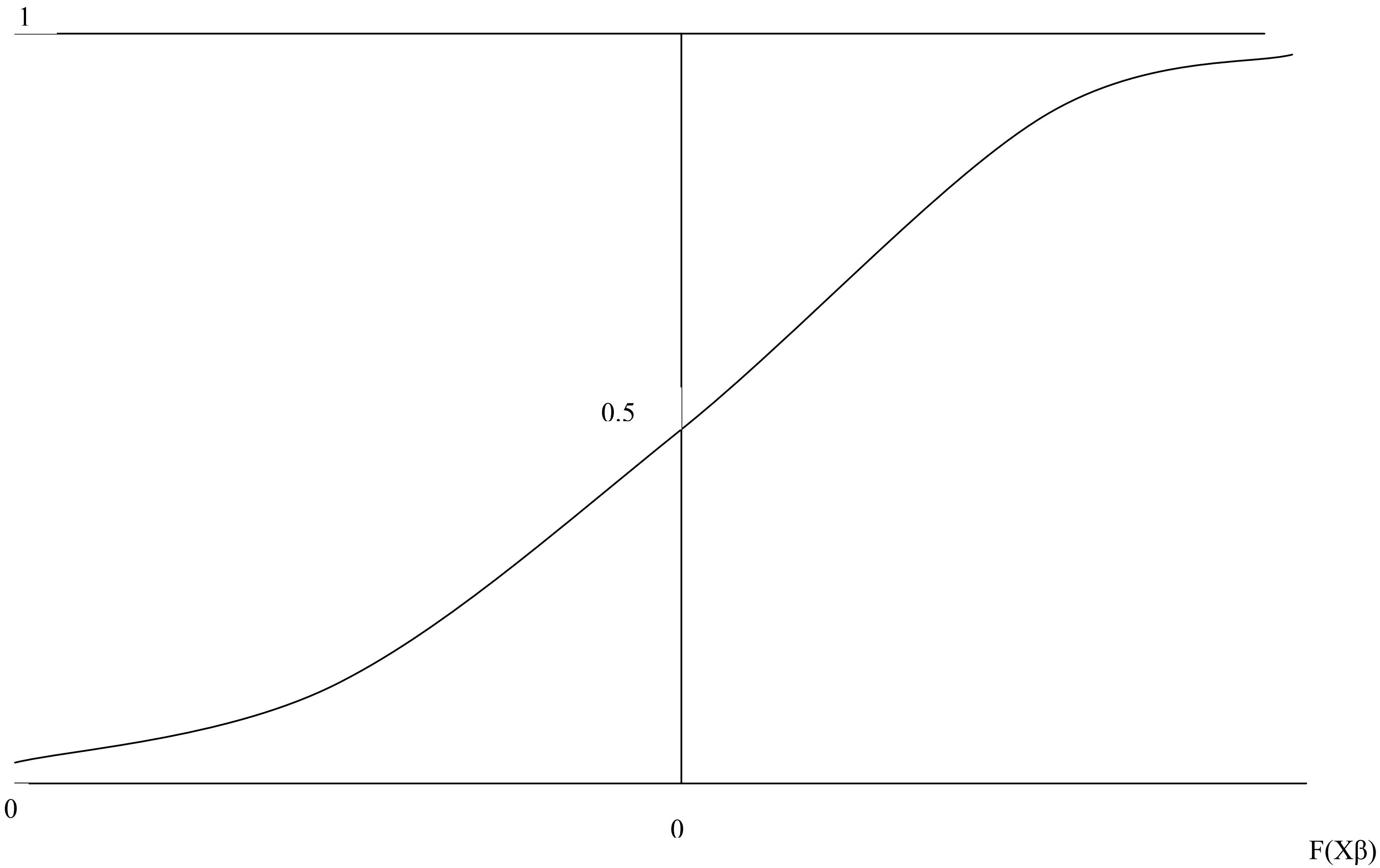
1. Logit Model

Assumes that the Probability model

$$\text{Probability } (D_i=1) = F(X_1, X_2 \ldots X_k)$$

is given by

$$\text{Prob } (D_i=1) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2)}}$$



a logit function is bounded by 0 and 1 and looks something like this – As the value of the logit function F(XB) rises the probability asymptotes to one. As the vale of F(XB) falls the probability asymptotes to zero

2. The Probit Model

Assumes that the Probability model is given by

$$\text{Prob } (D_i=1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\beta_0+\beta_1 X_1+\beta_2 X_2)^2}$$

The idea is to find the values of the coefficients β_0 , β_1 etc that make this probability as close to the values in the dataset as possible.

1. Logit Model
Assumes that the Probability model

$$\text{Probability } (D_i=1) = F(X_1, X_2 \ldots X_k)$$

is given by

$$\text{Prob } (D_i=1) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2)}}$$

2. The Probit Model
Assumes that the Probability model is given by

$$\text{Prob } (D_i=1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\beta_0+\beta_1 X_1+\beta_2 X_2)^2}$$

The idea is to find the values of the coefficients β_0 , β_1 etc that make this probability as close to the values in the dataset as possible.

The technique is called **maximum likelihood estimation**

Example: Using the marks.dta data set above the logit and probit equivalents of the OLS linear probability estimates above are, respectively

```
logit first num_sems female

Iteration 0:   log likelihood = -74.875501
Iteration 1:   log likelihood = -63.834471
Iteration 2:   log likelihood = -62.907479
Iteration 3:   log likelihood = -62.868479
Iteration 4:   log likelihood = -62.868381

Logistic regression                               Number of obs   =        118
                                                  LR chi2(2)      =      24.01
                                                  Prob > chi2     =     0.0000
Log likelihood = -62.868381                       Pseudo R2       =     0.1604
------------------------------------------------------------------------------
       first |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    num_sems |   .2397919   .0747599     3.21   0.001     .0932652    .3863185
      female |   1.054987   .4337783     2.43   0.015     .2047975    1.905177
       _cons |  -4.357462   1.084287    -4.02   0.000    -6.482626   -2.232298
------------------------------------------------------------------------------

probit first num_sems female

Iteration 0:   log likelihood = -74.875501
Iteration 1:   log likelihood = -63.637368
Iteration 2:   log likelihood = -63.127944
Iteration 3:   log likelihood = -63.122191
Iteration 4:   log likelihood =  -63.12219

Probit regression                                 Number of obs   =        118
                                                  LR chi2(2)      =      23.51
                                                  Prob > chi2     =     0.0000
Log likelihood =  -63.12219                       Pseudo R2       =     0.1570
------------------------------------------------------------------------------
       first |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    num_sems |   .1329781   .0397433     3.35   0.001     .0550827    .2108736
      female |   .6198004   .2607587     2.38   0.017     .1087228    1.130878
       _cons |   -2.44597   .5545824    -4.41   0.000    -3.532932   -1.359009
```

Note that the predicted values from the logit and probit regressions will lie between 0 and 1

```
predict phat_probit
(option p assumed; Pr(first))

su phat_probit phat_logit

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 phat_probit |       118    .3330643    .1980187   .0072231   .7147911
  phat_logit |       118    .3305085           0   .3305085   .3305085
```

while the means are the same the predictions are not identical for the two estimation techniques

The standard errors and t values on these variables should be free of the bias inherent in OLS – though they could still be subject to other types of heteroskedasticity so it is a good idea to use the “, robust” adjustment even with logit and probit estimators

```
logit first num_sems female, robust

Iteration 0:   log pseudolikelihood = -74.875501
Iteration 1:   log pseudolikelihood = -63.834471
Iteration 2:   log pseudolikelihood = -62.907479
Iteration 3:   log pseudolikelihood = -62.868479
Iteration 4:   log pseudolikelihood = -62.868381

Logistic regression                               Number of obs   =        118
                                                  Wald chi2(2)    =      14.65
                                                  Prob > chi2     =     0.0007
Log pseudolikelihood = -62.868381                 Pseudo R2       =     0.1604
------------------------------------------------------------------------------
             |               Robust
       first |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    num_sems |   .2397919   .0770628     3.11   0.002     .0887516    .3908322
      female |   1.054987   .4304162     2.45   0.014     .2113872    1.898588
       _cons |  -4.357462   1.154993    -3.77   0.000    -6.621206   -2.093718
------------------------------------------------------------------------------
```

(in this example it make little difference)

In both cases the estimated coefficients look very different from those of the OLS linear probability estimates.

In both cases the estimated coefficients look very different from those of the OLS linear probability estimates.
This is because they do not have the same interpretation as with OLS (they are simply values that maximise the likelihood function).

In both cases the estimated coefficients look very different from those of the OLS linear probability estimates.
This is because they do not have the same interpretation as with OLS (they are simply values that maximise the likelihood function).

To obtain coefficients which can be interpreted in a similar way to OLS, need **marginal effects**

In both cases the estimated coefficients look very different from those of the OLS linear probability estimates.
This is because they do not have the same interpretation as with OLS
(they are simply values that maximise the likelihood function).

To obtain coefficients which can be interpreted in a similar way to OLS, need **marginal effects**

$$\frac{\delta \Pr ob(D_i = 1)}{\delta X_i}$$

In both cases the estimated coefficients look very different from those of the OLS linear probability estimates.
This is because they do not have the same interpretation as with OLS (they are simply values that maximise the likelihood function).

To obtain coefficients which can be interpreted in a similar way to OLS, need **marginal effects**

$$\frac{\delta \Pr ob(D_i = 1)}{\delta X_i}$$

for the logit model this is given by

In both cases the estimated coefficients look very different from those of the OLS linear probability estimates.
This is because they do not have the same interpretation as with OLS
(they are simply values that maximise the likelihood function).

To obtain coefficients which can be interpreted in a similar way to OLS, need **marginal effects**

$$\frac{\delta \Pr ob(D_i = 1)}{\delta X_i}$$

for the logit model this is given by

$$\frac{\delta \Pr ob(D_i = 1)}{\delta X_i} = \beta_i \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2)}{(1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2))^2})$$

In both cases the estimated coefficients look very different from those of the OLS linear probability estimates.
This is because they do not have the same interpretation as with OLS (they are simply values that maximise the likelihood function).

To obtain coefficients which can be interpreted in a similar way to OLS, need **marginal effects**

$$\frac{\delta \Pr ob(D_i = 1)}{\delta X_i}$$

for the logit model this is given by

$$\frac{\delta \Pr ob(D_i = 1)}{\delta X_i} = \beta_i \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2)}{(1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2))^2})$$

$$= \beta_i \Pr ob(D_i = 1) * (1 - \Pr ob(D_i = 1))$$

In both cases the estimated coefficients look very different from those of the OLS linear probability estimates.
This is because they do not have the same interpretation as with OLS (they are simply values that maximise the likelihood function).

To obtain coefficients which can be interpreted in a similar way to OLS, need **marginal effects**

$$\frac{\delta \Pr ob(D_i = 1)}{\delta X_i}$$

for the logit model this is given by

$$\frac{\delta \Pr ob(D_i = 1)}{\delta X_i} = \beta_i \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2)}{(1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2))^2})$$

$$= \beta_i \Pr ob(D_i = 1) * (1 - \Pr ob(D_i = 1))$$

In both cases the estimated coefficients look very different from those of the OLS linear probability estimates.
This is because they do not have the same interpretation as with OLS
(they are simply values that maximise the likelihood function).

To obtain coefficients which can be interpreted in a similar way to OLS, need **marginal effects**

$$\frac{\delta \Pr ob(D_i = 1)}{\delta X_i}$$

for the logit model this is given by

$$\frac{\delta \Pr ob(D_i = 1)}{\delta X_i} = \beta_i \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2)}{(1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2))^2})$$

$$= \beta_i \Pr ob(D_i = 1) * (1 - \Pr ob(D_i = 1))$$

$$= \beta_i \Pr ob(D_i = 1) * (1 - \Pr ob(D_i = 1))$$

(in truth since Prob(D_i=1) varies with the values of the X variables, this marginal effect is typically evaluated with all the X variables set to their mean values)

In the case of probit model this marginal effect is given by

$$\frac{\delta \Pr ob(D_i = 1)}{\delta X_i} = \beta_i f\left(\beta_0 + \beta_1 X_1 + \beta_2 X_2\right)$$

(where f is the differential of the probit function – again typically evaluated at the mean of each X variable) and in the case of probit model this is given by

$$\frac{\delta \Pr ob(D_i = 1)}{\delta X_i} = \beta_i f\left(\beta_0 + \beta_1 X_1 + \beta_2 X_2\right)$$

(where f is the differential of the probit function above– again typically evaluated at the mean of each X variable)

In both cases the interpretation of these marginal effects is the impact that a unit change in the variable X_i has on the probability of belonging to the treatment group (just like OLS coefficients)

To obtain marginal effects in Stata run either the logit or probit command then simply type

```
logit first num_sems female
mfx

Marginal effects after logit
      y  = Pr(first) (predict)
         =  .27708689
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
num_sems |   .0480326      .01343    3.58   0.000   .021713  .074352   12.4576
  female*|   .2192534      .09097    2.41   0.016   .040952  .397554   .389831
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1


probit first num_sems female
mfx

Marginal effects after probit
      y  = Pr(first) (predict)
         =  .29192788
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
num_sems |   .0456601      .01301    3.51   0.000   .020157  .071163   12.4576
  female*|   .2177256      .09231    2.36   0.018   .036807  .398645   .389831
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1

(or with probit you can also type

dprobit first num_sems female

Iteration 0:   log likelihood = -74.875501
Iteration 1:   log likelihood = -63.637368
Iteration 2:   log likelihood = -63.127944
Iteration 3:   log likelihood = -63.122191
Iteration 4:   log likelihood =  -63.12219
```

```
Probit regression, reporting marginal effects           Number of obs =    118
                                                        LR chi2(2)    =  23.51
                                                        Prob > chi2   = 0.0000
Log likelihood =  -63.12219                             Pseudo R2     = 0.1570

------------------------------------------------------------------------------
   first |      dF/dx   Std. Err.      z    P>|z|     x-bar  [    95% C.I.   ]
---------+--------------------------------------------------------------------
num_sems |   .0456601   .0130119     3.35   0.001   12.4576   .020157  .071163
  female*|   .2177256   .0923073     2.38   0.017   .389831   .036807  .398645
---------+--------------------------------------------------------------------
  obs. P |   .3305085
 pred. P |   .2919279  (at x-bar)
------------------------------------------------------------------------------
(*) dF/dx is for discrete change of dummy variable from 0 to 1
    z and P>|z| correspond to the test of the underlying coefficient being 0
```

These estimates are similar to those of OLS (as they should be since OLS, logit and probit are unbiased)

Goodness of Fit Tests

The maximum likelihood equivalent to the F test of goodness of fit of the model as a whole is given by the **Likelihood Ratio (LR) test**

Goodness of Fit Tests

The maximum likelihood equivalent to the F test of goodness of fit of the model as a whole is given by the **Likelihood Ratio (LR) test**

which compares the value of the (log) likelihood when maximised with the likelihood function with all coefficients set to zero
(much like the F test compares the model RSS with the RSS when all coefficients set to zero)

Goodness of Fit Tests

The maximum likelihood equivalent to the F test of goodness of fit of the model as a whole is given by the **Likelihood Ratio (LR) test**

which compares the value of the (log) likelihood when maximised with the likelihood function with all coefficients set to zero
(much like the F test compares the model RSS with the RSS when all coefficients set to zero)

Can show that

$$LR = 2\,[\text{Log } L_{max} - \text{Log } L_0\,] \sim \chi^2_{(k-1)}$$

Goodness of Fit Tests

The maximum likelihood equivalent to the F test of goodness of fit of the model as a whole is given by the **Likelihood Ratio (LR) test**

which compares the value of the (log) likelihood when maximised with the likelihood function with all coefficients set to zero
(much like the F test compares the model RSS with the RSS when all coefficients set to zero)

Can show that

$$LR = 2\,[\text{Log } L_{max} – \text{Log } L_0\,] \sim \chi^2_{(k-1)}$$

where k is the number of right hand side variables including the constant

Goodness of Fit Tests

The maximum likelihood equivalent to the F test of goodness of fit of the model as a whole is given by the **Likelihood Ratio (LR) test**

which compares the value of the (log) likelihood when maximised with the likelihood function with all coefficients set to zero
(much like the F test compares the model RSS with the RSS when all coefficients set to zero)

Can show that

$$LR = 2\,[\text{Log } L_{max} - \text{Log } L_0\,] \sim \chi^2_{(k-1)}$$

where k is the number of right hand side variables including the constant

If the estimate chi-squared value exceeds the critical value then reject the null that the model as no explanatory power

Goodness of Fit Tests

The maximum likelihood equivalent to the F test of goodness of fit of the model as a whole is given by the **Likelihood Ratio (LR) test**

which compares the value of the (log) likelihood when maximised with the likelihood function with all coefficients set to zero
(much like the F test compares the model RSS with the RSS when all coefficients set to zero)

Can show that

$$LR = 2\,[\text{Log } L_{max} - \text{Log } L_0\,] \sim \chi^2_{(k-1)}$$

where k is the number of right hand side variables including the constant

If the estimate chi-squared value exceeds the critical value then reject the null that the model as no explanatory power
(this value is given in the top right hand corner of the logit/probit output in Stata)

Can also use the LR test to test restrictions on subsets of the coefficients in a similar way to the F test

$$LR = 2\,[\text{Log } L^{max}_{unrestrict} - \text{Log } L^{max}_{restrict}\,] \sim \chi^2_{(l)}$$

(where l is the number of restricted variables)

A maximum likelihood equivalent of the R^2 is the **pseudo-R^2**

A maximum likelihood equivalent of the R^2 is the **pseudo-R^2**

$$\text{Pseudo } R^2 = 1 - (\text{Log } L_{max}/\text{Log } L_0)$$

A maximum likelihood equivalent of the R^2 is the **pseudo-R^2**

Pseudo $R^2 = 1 – (\text{Log } L_{max}/\text{Log } L_0)$

This value lies between 0 and 1 and the closer to one the better the fit of the model

(again this value is given in the top right hand corner of the logit/probit output in Stata)

It is also a good idea to try and look at the “% of correct predictions” in the model – ie how many are predicted to have a value 1 and how many predicted to have a value 0

Can do this by assigning a rule

Predict = 1 if $\hat{p} >= .5$

Predict = 0 if $\hat{p} < .5$

where $\hat{p}$ is the individual predicted probability taken from the logit or probit model

```
g predict=1 if phat>=.5
replace predict=0 if phat<.5

tab predict first, row
           |        first
   predict |         0          1 |     Total
-----------+----------------------+----------
         0 |        69         22 |        91
           |     75.82      24.18 |    100.00
-----------+----------------------+----------
         1 |        10         17 |        27
           |     37.04      62.96 |    100.00
-----------+----------------------+----------
     Total |        79         39 |       118
           |     66.95      33.05 |    100.00
```

So in this case the model predicts 63% of firsts correctly and 75% of non-firsts correct. Compare this with a random guess which would get 50% of each category correct