Acknowledgement of Country

I acknowledge the Traditional Owners and their custodianship of the lands on which we meet today.

On behalf of us all, I pay our respects to their Ancestors and their descendants, who continue cultural and spiritual connections to Country.

We recognise their valuable contributions to Australian and global society.

Image: Digital reproduction of *A guidance through time* by Casey Coolwell and Kyra Mancktelow

Lecture 11: Limited dependent variable models

Julie Moschion
University of Queensland

Reading for Lecture 11

- In the textbook (Wooldridge 2020): Chapters 7 and 17

Introduction

- Up until now, we have modelled the relationship between the explanatory variables and the dependent variable using the multiple regression model:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u$$

- This assumes that the dependent variable is not limited and can take any value: negative values, decimal values...
- However, often we want to explain an outcome that is limited in some ways:
  - maybe it only takes a few values (e.g. the number of children in a family),
  - maybe even just 0 or 1 (e.g. whether a family had another child in year t),
  - or maybe most of the population bunches at certain values of the outcome (e.g. in a young population most people will have 0 children).

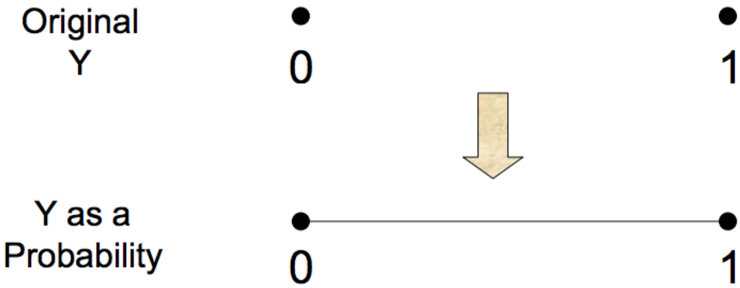
Introduction

- It is often ok to use the multiple regression model even when the outcome is only approximately continuous.
- But there are times when there is a better model available, or at least an alternative way to estimate the relationship of interest...
- In this lecture, we focus on **binary outcomes, i.e. outcomes that only take two values: 0 or 1**. we also refer to these as: "dummy dependent variable".
- These outcomes are extremely common in the real world and therefore the question of the model used very common in applied economics!

Introduction

- A few examples which can all be coded as (y=1 if yes; 0 if not) include:
  - Voting,
  - Being married,
  - Having Children,
  - Being pro-immigration
  - Exercising regularly
  - Succeeding a test
  - Taking illegal drugs...
- When the dependent variable is a binary, we should be thinking in terms of **probabilities**, i.e. a number between 0 and 1.

Introduction



Introduction

- Instead of a continuous y, we have to model the probability that y=0 or y=1, given the explanatory variables x:

$$P = P(y = 1|x) = F(x'\beta) = p(x)$$

- P is called the response probability: it is the probability of a "success", i.e. $y = 1$.
- Of course:
$P(y = 0|x) = 1 - P(y = 1|x) = 1 - F(x'\beta) = 1 - p(x)$
- As in multiple linear regressions, we are interested in the partial effects of the x_j on the outcome, here: $p(x)$, i.e. by how much does the outcome vary when x_j varies by 1 unit.

The 3 models for binary outcomes

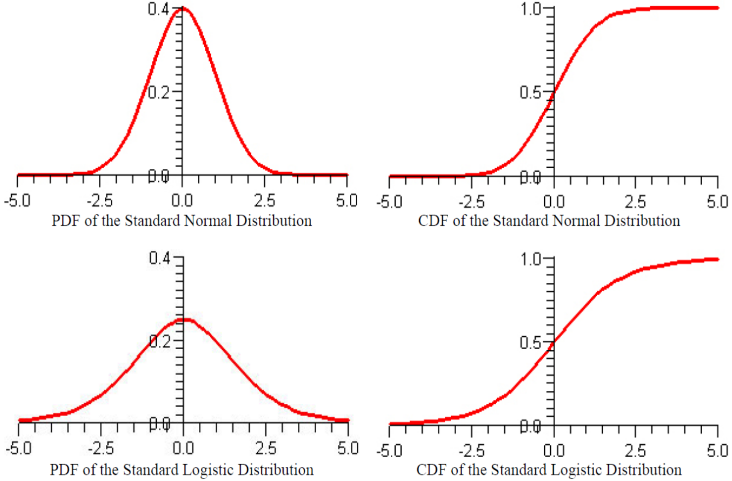
- There are 3 models that handle binary outcomes: Linear Probability Models (LPM), Probit and Logit models:
  - **LPMs** have the simplest functional form with: $F(x'\beta)=x'\beta$. LPMs are estimated via OLS.
  - **Logit and probit models** have more complex functional forms and are estimated via Maximum Likelihood (ML) because of their non-linear nature.

The 3 models for binary outcomes

- **The difference between logit and probit models lies in the assumptions about the distribution of the errors.**
- For the logit model:
  - The errors are assumed to follow a logistic distribution.
  - The functional form is the cumulative distribution function (cdf) for the logistic regression: $F(x'\beta) = \frac{e^{x'\beta}}{1+e^{x'\beta}}$.
- For the probit model:
  - The errors are assumed to follow a normal distribution.
  - The functional form is the cdf for the probit regression (the cdf of the normal distribution): $F(x'\beta) = \Phi(x'\beta) = \int_{-\infty}^{\mathbf{x}'\beta} \phi(z)dz$.
- The logistic and normal distribution have the same shape but the logistic is flatter on the top and fatter in the tails.

The 3 models for binary outcomes

Probability density function (PDF) & cumulative distribution function (CDF)



The 3 models for binary outcomes

- **How does one choose between the 3 models?**
  - Results tend to be very similar.
  - Preference for one over the other tends to vary by discipline.
  - Applied economists might have a slight preference for LPM because the interpretation of coefficients is easier but LPM has some issues (discussed a bit later in the class).
- Usually with binary outcomes, results are presented for the LPM and either a logit or probit with one being the main specification and the other a robustness check. The choice between which is the main and which is a robustness is your choice!

The linear probability model (LPM)

- LPMs are nice because estimation is easy (it's a simple OLS) and so is interpretation.
- The LPM is a standard linear model where y happens to be binary:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u$$

- How do we interpret the coefficients when y is binary?
  - Suppose y denotes being married or not: y can only change from 0 to 1 or 1 to 0.
  - Suppose β_1=0.03 and $x_1 = educ$.
  - What does it mean that an extra year of education is associated with an increase in marital status by 0.03?

The linear probability model (LPM)

- Remember we can also write the regression model as a linear response probability model (hence the name LPM):
$$\begin{aligned} P(y = 1|X) &= \beta_0 + \beta_1 x_1 + ... + \beta_k x_k \\ &= E(y|X) \end{aligned}$$

- we can interpret β_j as

$$\Delta E(y|X) = \beta_j \Delta x_j$$
$$i.e.\ \beta_j = \frac{\Delta E(y|X)}{\Delta x_j}$$

while holding other explanatory variables constant.

The linear probability model (LPM)

- $\hat{\beta}_j$ measures the change in the estimated probability of a "success" when $\Delta x_j = 1$ (other factors held constant).
- In the above example: **if education increases by 1 year then the individual is 3 percentage points (often noted pp) more likely to be married.**
- Importantly the increase is 3pp, not 3%.
  - Suppose that 55% of individuals with 8 years of education are married. A coefficient of 0.03 means that for individuals with 9 years of education, the probability to be married is 58%.
  - This is not a 3% increase but a 5% increase: $(((58/55) - 1) * 100)$.
- Note that explanatory variables can be in logs, quadratics, and interactions as well as binary regressors.

LPM example: mortgage application

- When people want to buy a house, they often apply for a mortgage at a bank.
- The outcome of this application is a dummy outcome: it is either successful or not.
- Each mortgage application has a certain probability of success given its observed characteristics (X).
- The model relating these characteristics (X) to the outcome (y) of the mortgage application can be written as a LPM.
- What characteristics may determine whether a mortgage application is approved or denied?
  - Often banks consider that 33% max of the household income should be spent on repayments.
  - Among others, the ratio of repayments to income is a characteristic that will affect whether a loan is approved or not.

LPM example: mortgage application

- Let's run a regression of the probability that the mortgage is denied on the payment to income ratio. Note that the outcome is defined negatively: that's ok p(approved)=1-p(denied).

```
. regress deny pi_ratio, robust

Linear regression                                      Number of obs =    2380
                                                         F(  1,  2378) =   37.56
                                                         Prob > F      =  0.0000
                                                         R-squared     =  0.0397
                                                         Root MSE      =  .31828

------------------------------------------------------------------------------
             |               Robust
        deny |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    pi_ratio |   .6035349   .0984826     6.13   0.000     .4104144    .7966555
       _cons |  -.0799096   .0319666    -2.50   0.012    -.1425949   -.0172243
------------------------------------------------------------------------------
```

- The OLS coefficient is 0.60: a change in the payment to income ratio from 0 to 1 (i.e. from 0% to 100%) increases the probability that the mortgage is denied by 60pp.
- **A change in the payment to income ratio by 10pp (e.g. from 30% to 40%) increases the probability to be denied by 6pp** (0.10*0.60*100).

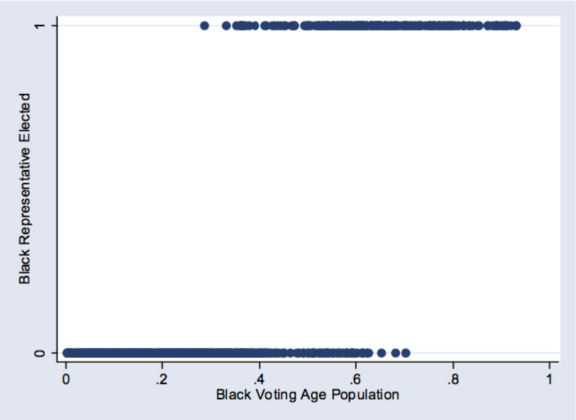
Advantages of LPM

- LPMs are linear models, so they are simple to estimate.
- They are also simple to interpret: the estimated coefficients $\hat{\beta}_j$ directly give the marginal effects (i.e. by how much y changes when x_j changes by 1 unit).
- Inference is the same as for multiple regression models.

Problems with the LPM

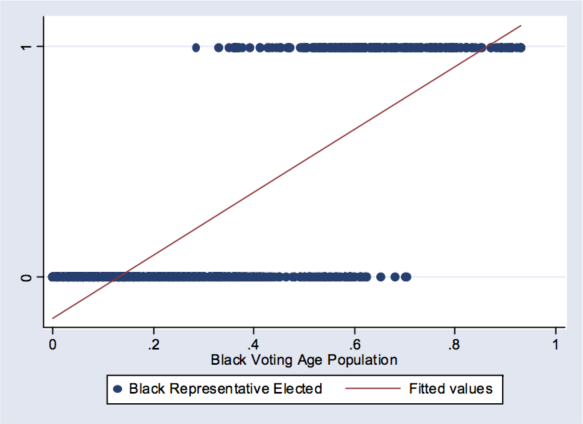
- But LPMs also have problems!
- Imagine that you want to study the election of minorities in the U.S (State of Georgia).
- To do so, you estimate a LPM model in which the outcome $y = 1$ if the elected person is from a minority group and $y = 0$ if the elected person is not from a minority group.
- In this model, one of the characteristics that may influence the outcome of the election is the percentage of black people in the voting-age population (x).
- Let's plot y and x from data in a number of local elections.

Problems with the LPM



- x varies between 0 and 1, but y is either 0 or 1.
- This looks a bit odd compared to the cloud of dots you usually get, right?

Problems with the LPM



- We add a linear fit, which is what a linear regression does. The slope corresponds to the difference between the proportion of localities that elect a minority and those that don't, averaged across all values of x.
- But the line doesn't fit the data very well as it appears possibly non-linear i.e. not all increases in x result in the same increase in y.
- Also, note that the line goes below 0 and above 1.

Problems with the LPM

- Sometimes we also use LPM to predict probabilities.
- Imagine you are a married man in the 1970s and you suspect your wife of having an affair.
- You have data from a survey conducted by Psychology Today in 1969 on a sample of 601 readers: men and women, employed and in their first marriage.
- You want to use these data to assess the likelihood that your wife is cheating on you given her characteristics.

Problems with the LPM

- Given the information available, you set out to estimate the following model:

$$Affair = \beta_0 + \beta_1 Male + \beta_2 Age + \beta_3 YearsMarried + \beta_4 WithKids + \beta_5 Relig + \beta_6 YrsEducation + \beta_7 Ratemarr + u$$

- Where:
  - $Affair = 1$ if the respondent had at least 1 affair, 0 otherwise;
  - *Relig* is a categorical variable ranging from low religiosity (1) to high (5);
  - *Ratemarr* is a categorical variable ranging from low happiness in the marriage (1) to high (5).

Problems with the LPM

```
. reg affair male age yrsmarr kids relig educ ratemarr, r
```

Linear regression

Number of obs	=	601
F(7, 593)	=	9.61
Prob > F	=	0.0000
R-squared	=	0.1062
Root MSE	=	.41189

affair	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
male	.0517169	.0389454	1.33	0.185	-.0247707	.1282046
age	-.0073149	.0031354	-2.33	0.020	-.0134728	-.0011571
yrsmarr	.0160812	.00563	2.86	0.004	.005024	.0271384
kids	.050184	.0449966	1.12	0.265	-.0381881	.1385561
relig	-.0539739	.0153308	-3.52	0.000	-.0840832	-.0238647
educ	.004867	.0078468	0.62	0.535	-.0105439	.020278
ratemarr	-.087728	.0170375	-5.15	0.000	-.1211892	-.0542668
_cons	.729658	.161796	4.51	0.000	.4118952	1.047421

Problems with the LPM

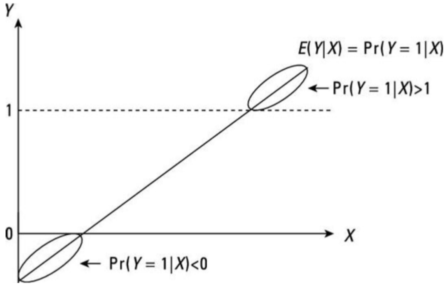
- Your wife is a very religious 25-year-old woman with 2 years of education, no kids and very happily married for 1 year.
- What is the expected probability that she is having an affair?

$$\begin{aligned} P(\widehat{Affair = 1}) &= 0.73 + 0.051 Male - 0.007 Age \\ &+ 0.016 YearsMarried + 0.05 WithKids \\ &- 0.054 Relig + 0.05 YrsEducation - 0.088 Ratemarr \\ &= 0.73 - 0.007 * 25 + 0.016 - 0.054 * 5 \\ &+ 0.05 * 2 - 0.088 * 5 \\ &= -0.04 \end{aligned}$$

- But a probability cannot be negative!
- LPMs can predict probabilities outside of [0,1] though.

Problems with the LPM

- There are 3 problems with the LPM
  - **Unbounded predicted probabilities**: probabilities are within the interval [0,1]. But LPMs do not ensure that this is satisfied. Although most of the predicted probabilities from a LPM fall between 0 and 1, some predicted probabilities can be less than 0 or greater than 1. This is a bit awkward but not a big deal in practice: we mostly use LPMs to estimate marginal effects, not to make predictions.



Problems with the LPM

- **Heteroskedasticity**: the error does not have the same variance for any values of the Xs. This violates one of the Gauss-Markov assumptions (A5) such that estimates from the LPM are not BLUE (best linear unbiased estimator). But like we have seen before, this is easily corrected with robust standard errors.
- **Non-Normality**: the distribution of the error term is not normal (it can only take 2 values: 0 and 1) so A6 which governs calculations of t-stats and F-stats is violated.

- Importantly, despite these issues **LPM estimates remain unbiased** estimates of the true parameter β (A1-A4 i.e. linearity, random sample, non collinearity of Xs and exogeneity hold).

Problems with the LPM

- Given its simplicity and unbiased properties, many researchers favour the LPM. A prominent example is Joshua Angrist (2009):
  - "...while a nonlinear model may **fit** ... more closely than a linear model, when it comes to marginal effects, this probably matters little... Nonlinear life also gets considerably more complicated when we work with instrumental variables and panel data."
- And in practice the marginal effects from the LPM are very close to the Probit or the Logit estimates (see, for example, Wooldridge, 2002 and Angrist and Pischke, 2009).

Probit and logit models

- Probit and logit models are estimated via Maximum Likelihood Estimator (MLE). The basic idea is the following:
  - You have a likelihood function based on the joint distribution of the data (y_i, x_i) for $i = 1, 2, 3, ..., n$ and the unknown coefficients (β).
  - The idea is to choose the values of β that maximise the fit of the model to the data by:
    - Taking the logarithm of the likelihood function;
    - Maximising this log likelihood function by taking the derivative with respect to β;
    - Solving for β.

Probit and logit models

- First let's consider the simple example of the Bernoulli disbribution.
- Consider an experiment in which there are only two possible outcomes: success (value of 1) or failure (value of 0).
- The random variable y has a Bernoulli distribution with parameter p if:

$$P(Y = 1) = p$$
$$P(Y = 0) = 1 - p$$

where success occurs with probability p and failure occurs with probability 1-p. In this simple case p is set (it is not a function of X and β).

Probit and logit models

- The likelihood of the Bernoulli distribution is:

$$L = \prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i}$$

- The log-likelihood is:

$$lnL = \sum_{i=1}^{n} y_i ln(p) + (1-y_i) ln(1-p)$$

- The MLE finds the p that maximises this log likelihood function by taking the derivative with respect to p and solving.

Probit and logit models

- But the Bernoulli distribution is only an appropriate model if each event had the same chance of occurring, i.e. when p completely describes y and is independent of X and β. For example, if you flip a coin, the probability to get tails is 0.5.
- However, it is not a good model when the realisation of the outcome depends on a number of characteristics.
- For example, not all voters have the same probability of voting or all countries of going to war.

Probit and logit models

- A more general model that ensures probabilities are between 0 and 1 can be written as:
$$P(y = 1|X) = F(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k)$$
- $F(\cdot)$ is a cumulative distribution function (CDF) for a continuous random variable where the area under the CDF curve to the left of $x'\beta$ gives the cumulative probability of y given $x'\beta$.
- For the logit model:
  - The CDF is $\mathrm{F}(\mathrm{x}'\beta) = \frac{e^{x'\beta}}{1+e^{x'\beta}} = \Lambda\ (\mathrm{x}'\beta)$.
- For the probit model (normal distribution):
  - The CDF is $\mathrm{F}(\mathrm{x}'\beta) = \int_{-\infty}^{\mathbf{x}'\beta} \phi(z)dz = \Phi(x'\beta)$.

Probit and logit models

- Probit likelihood & log-likelihood functions

$$L(\beta|X) = \prod_{i=1}^{n}[(1-\Phi(X'\beta))^{1-y_i} \cdot \Phi(X'\beta)^{y_i}]$$
$$\ln L = \Sigma_{i=1}^{n}(1-y_i) \cdot \ln(1-\Phi(X'\beta)) + \Sigma_{i=1}^{n} y_i \cdot ln\Phi(X'\beta)$$

- Logit likelihood & log-likelihood functions

$$L(\beta|X) = \prod_{i=1}^{n}[(1-\Lambda(X'\beta))^{1-y_i} \cdot \Lambda(X'\beta)^{y_i}]$$
$$\ln L = \Sigma_{i=1}^{n}(1-y_i) \cdot \ln(1-\Lambda(X'\beta)) + \Sigma_{i=1}^{n} y_i \cdot ln\Lambda(X'\beta)$$

- And then maximising this log likelihood function by taking the derivative with respect to β and solving for β.

Advantages and disadvantages

- The clear advantage is that the predicted probabilities are bounded between 0 and 1.
- But the disadvantage is that the coefficients are difficult to interpret. So marginal effects will need to be calculated post-estimation to interpret estimated effects.

Probit example: mortgage application

- Let's go back to our mortgage example and estimate the model using a probit.

```
. probit deny pi_ratio

Iteration 0:   log likelihood =  -872.0853
Iteration 1:   log likelihood = -832.02975
Iteration 2:   log likelihood = -831.79239
Iteration 3:   log likelihood = -831.79234

Probit regression                               Number of obs   =       2380
                                                LR chi2(    1)      =      80.59
                                                Prob > chi2     =     0.0000
Log likelihood =  -831.79234                     Pseudo R2       =     0.0462

------------------------------------------------------------------------------
        deny |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    pi_ratio |   2.967907   .3591054     8.26   0.000     2.264073     3.67174
       _cons |  -2.194159     .12899   -17.01   0.000    -2.446974   -1.941343
------------------------------------------------------------------------------
```

- The estimated MLE coefficient on the payment to income ratio is 2.97 (the OLS coefficient was 0.60).
- It is positive and significant at 1%.
- But what does it mean: how do we interpret its magnitude?

Probit example: mortgage application

- One way to interpret the estimates is to calculate the difference in the predicted outcome (probability that the mortgage is denied) when the explanatory variable (PI ratio) increases from a specified value to another value:
  - PI increases from 10% to 20%:

$$\Delta \widehat{P(Y=1)} = \Phi(-2.19 + 2.97 * 0.2) - \Phi(-2.19 + 2.97 * 0.1)$$
$$= 0.0495$$

  - PI increases from 30% to 40%:

$$\Delta \widehat{P(Y=1)} = \Phi(-2.19 + 2.97 * 0.4) - \Phi(-2.19 + 2.97 * 0.3)$$
$$= 0.0619$$

- Note the non-linearity of the effects: **a change in the payment to income ratio from 10% to 20% (resp. from 30% to 40%) increases the probability to be denied by 5pp (resp. 6pp)**.
- This is very close to the LPM estimate of 6pp.

Logit example: mortgage application

- Let's now estimate the model using a logit.

```
. logit deny pi_ratio

Iteration 0:   log likelihood =   -872.0853
Iteration 1:   log likelihood =  -830.96071
Iteration 2:   log likelihood =  -830.09497
Iteration 3:   log likelihood =  -830.09403
Iteration 4:   log likelihood =  -830.09403

Logistic regression                               Number of obs   =       2380
                                                  LR chi2(   1)      =      83.98
                                                  Prob > chi2     =     0.0000
Log likelihood =  -830.09403                       Pseudo R2       =     0.0482

------------------------------------------------------------------------------
        deny |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    pi_ratio |   5.884498   .7336006     8.02   0.000     4.446667    7.322328
       _cons |  -4.028432   .2685763   -15.00   0.000    -4.554832   -3.502032
------------------------------------------------------------------------------
```

- The estimated MLE coefficient on the payment to income ratio is 5.88.
- It is positive and significant at 1%.
- But what does it mean: how do we interpret its magnitude?

Logit example: mortgage application

- We can calculate the same predicted change in the outcome when the PI ratio increases:
  - PI increases from 10% to 20%:

$$\Delta \widehat{P(Y=1)} = \frac{e^{-4.03+5.88*0.2}}{1+e^{-4.03+5.88*0.2}} - \frac{e^{-4.03+5.88*0.1}}{1+e^{-4.03+5.88*0.1}}$$
$$= 0.023$$

  - PI increases from 30% to 40%:

$$\Delta \widehat{P(Y=1)} = \frac{e^{-4.03+5.88*0.4}}{1+e^{-4.03+5.88*0.4}} - \frac{e^{-4.03+5.88*0.3}}{1+e^{-4.03+5.88*0.3}}$$
$$= 0.063$$

- **A change in the payment to income ratio from 10% to 20% (resp. from 30% to 40%) increases the probability to be denied by 2pp (resp. 6pp)**.
- This is again very close to the LPM estimate of 6pp.

Marginal Effects

- But what about the overall average effect? This is what we call the marginal effect.
- In this example, we are interested in the probability of arrest for male adults as a function of their income (in $100) and minority status.

```
. logit arrest minority inc86

Logistic regression                               Number of obs   =       2725
                                                  LR chi2(2)      =     152.22
                                                  Prob > chi2     =     0.0000
Log likelihood = -1532.0747                       Pseudo R2       =     0.0473

------------------------------------------------------------------------------
      arrest |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    minority |   .5853512   .0886866     6.60   0.000     .4115286    .7591738
       inc86 |  -.0074475   .0008404    -8.86   0.000    -.0090947   -.0058003
       _cons |  -.8499352    .069239   -12.28   0.000    -.9856411   -.7142294
------------------------------------------------------------------------------
```

- The likelihood of being arrested is higher for minorities and decreases with income.
- But how large are these effects?

Marginal Effects

- To get the sense of the magnitude, we can calculate marginal effects after the logit estimation:

```
. mfx

Marginal effects after logit
      y  = Pr(arrest) (predict)
         =  .26160966
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
minority*|   .1165188        .018    6.47   0.000   .081238    .1518   .378716
   inc86 |  -.0014386      .00016   -9.17   0.000  -.001746 -.001131    54.967
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

- Q: For males with the average level of income in this sample, how much more likely are minorities to be arrested?
- A: 11.7 percentage points.

Marginal Effects

```
use http://dss.princeton.edu/training/Panel101.dta
quietly logit y_bin x1 x2 x3 i.opinion
margins, dydx(*) atmeans post
```

```
. margins, dydx(*) atmeans

Conditional marginal effects                    Number of obs   =         70
Model VCE    : OIM

Expression   : Pr(y_bin), predict()
dy/dx w.r.t. : x1 x2 x3 2.opinion 3.opinion 4.opinion
at           : x1              =    .6480006 (mean)
               x2              =    .1338694 (mean)
               x3              =     .761851 (mean)
               1.opinion       =    .2857143 (mean)
               2.opinion       =    .2142857 (mean)
               3.opinion       =    .2714286 (mean)
               4.opinion       =    .2285714 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .1384634   .1093955     1.27   0.206    -.0759478    .3528746
          x2 |    .036904   .0421082     0.88   0.381    -.0456266    .1194346
          x3 |     .04857   .0548416     0.89   0.376    -.0589176    .1560577
             |
     opinion |
       Agree |  -.3656898   .1670551    -2.19   0.029    -.6931118   -.0382678
       Disag |   .0312784   .0945857     0.33   0.741    -.1541062    .2166629
   Str disag |   .0574484    .098205     0.58   0.559    -.1350299    .2499268
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

Marginal effects show the change in probability when the predictor or independent variable increases by one unit. For continuous variables this represents the instantaneous change given that the 'unit' may be very small. For binary variables, the change is from 0 to 1, so one 'unit' as it is usually thought.

The change in probability for one instant change in *x1* is 13 percentage points (pp), in *x2* is 3 pp and in *x3* is 4 pp.
None of the effects here are significant (see column P>|z|, for significance at 95% values should be < 0.05)

1. The change in probability when opinion goes from 'strongly agree' to 'agree' decreases 36 percentage points or -0.36, and is significant.
2. The change in probability when opinion goes from 'strongly agree' to 'disagree' increases by 3 percentage points or 0.03.
3. The change in probability when opinion goes from 'strongly agree' to 'strongly disagree' increases by 5 percentage points or 0.05.

Type `help margins` for more details

In Stata

- To estimate a probit: `probit` y x_1 x_2 `...` x_k
- To estimate a logit: `logit` y x_1 x_2 `...` x_k
- Post-estimation, to calculate marginal effects: `margins, dydx(*) atmeans`. This works in the same way for logits and probits.

Multinomial choice models

- Sometimes the outcome of interest has more than a 0 or 1 option.
- For example, it can have multiple alternatives such as:
  travel modes (bus/train/car),
  employment status (employed/unemployed/retired),
  car choices (sedan/ute/convertible/minivan)
  ...and there is no particular meaning to the ordering.
- Usually, the aim is to estimate the impact of changes in individual characteristics (e.g. education, income, firm size etc.) on the probability of choosing an alternative outcome.

Multinomial choice models

- For example how does the probability of being employed (vs unemployed vs retired) differ for individuals who completed high school compared with those who dropped out?
- In these case, the appropriate models are multinomial probit and logit models. They are readily available in Stata.
- Note that in practice researchers often come back to a dummy option as it is simpler to interpret: for example the probability of being employed vs not being employed (rather than against each other option).