Introduction to Logit

Categorical and Limited Dependent Variables

Paul A. Jargowsky

Binary Dependent Variables

If everybody has the same probability:

$$Y_i = \begin{cases} 1 & \text{event occurs} \\ 0 & \text{does not occur} \end{cases} \qquad P = \text{ probability event occurs}$$

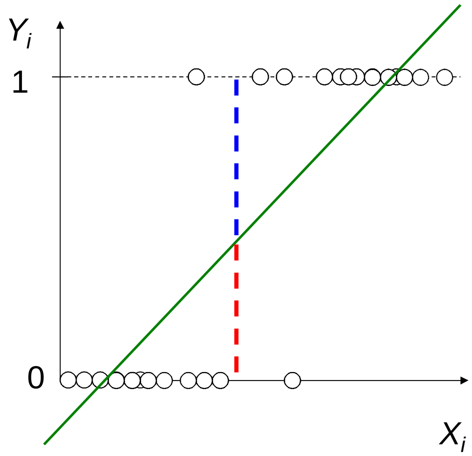
$$Y_i = \beta_1 + u_i \qquad \hat{\beta}_1 = \bar{Y} - 0\bar{X} = \bar{Y} \qquad \bar{Y} = \frac{\sum Y_i}{N} = P$$

If probability varies by individual according to X:

$$Y_i = \begin{cases} 1 & \text{event occurs} \\ 0 & \text{does not occur} \end{cases} \qquad P_i = \text{ probability event occurs to person } i$$

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_i + u_i \\ &= E(Y_i \mid X_i) + u_i \end{aligned} \qquad \begin{aligned} E\left[Y_i \mid X_i\right] &= P_i(1) + \left(1 - P_i\right)\left(0\right) \\ &= P_i \end{aligned}$$

Linear Probability Model



$$\hat{P}_i = \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

$$e_i = \begin{cases} 1 - \hat{P}_i \text{ if } Y_i = 1 \\ 0 - \hat{P}_i \text{ if } Y_i = 0 \end{cases}$$

$$\frac{d\hat{P}_i}{dX_i} = \hat{\beta}_2$$

Problems with LPM

- Nonsense predictions
  - The predicted conditional probabilities can be greater than 1 or less than 0.
- Heteroskedasticity
  - $Var(u_i)=P(1-P)$, $P=f(X)$
- Non-normal errors
  - Only two possible values
- Wrong functional form
  - Constant rate of change

Solution: choose a convenient functional form

- Find a transformation of P that can be modeled as linear

$$T(P_i) = \beta_1 + \beta_2 X_i + \ldots + \beta_K X_K$$

- This is only one way of deriving the logit model. We will talk about others later.
- The transformation should result in a smoothly varying, unbounded RV that can be modeled with a linear function

P	1-P	$\frac{P}{1-P}$	$\ln\left(\frac{P}{1-P}\right)$
===	===	====	======
0.001	0.999	0.00100	-6.90675
0.01	0.99	0.01010	-4.59512
0.1	0.9	0.11111	-2.19722
0.2	0.8	0.25000	-1.38629
0.3	0.7	0.42857	-0.84730
0.4	0.6	0.66667	-0.40547
0.5	0.5	1.00000	0.00000
0.6	0.4	1.50000	0.40547
0.7	0.3	2.33333	0.84730
0.8	0.2	4.00000	1.38629
0.9	0.1	9.00000	2.19722
0.99	0.01	99.00000	4.59512
0.999	0.001	999.00000	6.90675

We are looking for an unbounded, continuous transformation of P

The log of the odds is unbounded, continuous, and symmetrical and will serve our purposes well.

The Logit as a DV

$$\ln\left(\frac{P_i}{1-P_i}\right) = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_K X_{Ki} + u_i = \mathbf{x}_i\boldsymbol{\beta}$$

- The left side is known as the logit transformation.
- In grouped data, the percentage in each group can be a DV in an OLS regression, if none of the aggregated units has P=0 or P=1.
  - But do you want to use grouped data in your analysis? Why or why not?
- In individual level data, the DV would always be missing, because the “P” is just 0 or 1 for an individual, and the logit is undefined for those values.

Quick Math Review: Logarithms

$2^3 = ?$ What is 2 raised to the power of 3? Answer: 8.

$2^? = 8$ To get 8, what power would 2 have to be raised to? Answer: 3.

$base^{exponent} = result \qquad \log_{base}(result) = exponent$

$2^? = 8 \quad \rightarrow \quad \log_2 8 = 3$

Raising a number to a power (exponentiation) is the inverse function of taking a logarithm:

$$\log_2\left(2^3\right) = \log_2(8) = 3$$
$$2^{\log_2(8)} = 2^3 = 8$$

Natural Logarithms

Any number can serve as the base of logarithms. It doesn't matter what base you use, the properties are the same.
The most common basis for logarithms are 2, 10, and e = 2.712828….

When you see log(*x*) with no base shown, it ususally means the logarithm to base 10. When e is used, the symbol is usually ln(*x*), but it works the same way.

$\ln(8) \approx 2.08 \qquad e^{2.08} = 2.71828^{2.08} \approx 8.00$

```
. display ln(8)
2.0794415
```

```
. display exp(2.0794415)
7.9999997
```

```
. disp 2.71828^2.0794415
7.9999885
```

Useful Properties of Logarithms

$$\ln(xy) = \ln(x) + \ln(y)$$
$$\ln\left(\frac{x}{y}\right) = \ln(x) - \ln(y)$$
$$\ln\left(x^y\right) = y\ln(x)$$

Logarithms turn multiplication and division into addition and subtraction.
Also turn exponents into multiplication.

$$Y_i = aX_i^b u_i \quad \rightarrow \quad \ln Y_i = \ln a + b\ln X_i + \ln u_i$$
$$= \beta_1 + \beta_2 \ln X_i + \tilde{u}_i$$

For MLE, develop an expression for P_i

$$\ln\left(\frac{P_i}{1-P_i}\right) = \mathbf{x}_i\boldsymbol{\beta}$$

Solve for P_i:

$$e^{\left[\ln\left(\frac{P_i}{1-P_i}\right)\right]} = e^{\mathbf{x}_i\boldsymbol{\beta}} \quad \rightarrow \quad \frac{P_i}{1-P_i} = e^{\mathbf{x}_i\boldsymbol{\beta}}$$

$$P_i = e^{\mathbf{x}_i\boldsymbol{\beta}}\left(1-P_i\right) = e^{\mathbf{x}_i\boldsymbol{\beta}} - P_i e^{\mathbf{x}_i\boldsymbol{\beta}}$$

$$P_i + P_i e^{\mathbf{x}_i\boldsymbol{\beta}} = e^{\mathbf{x}_i\boldsymbol{\beta}}$$

$$P_i\left(1+e^{\mathbf{x}_i\boldsymbol{\beta}}\right) = e^{\mathbf{x}_i\boldsymbol{\beta}}$$

$$\boxed{P_i = \frac{e^{\mathbf{x}_i\boldsymbol{\beta}}}{1+e^{\mathbf{x}_i\boldsymbol{\beta}}}}$$

What happens when:

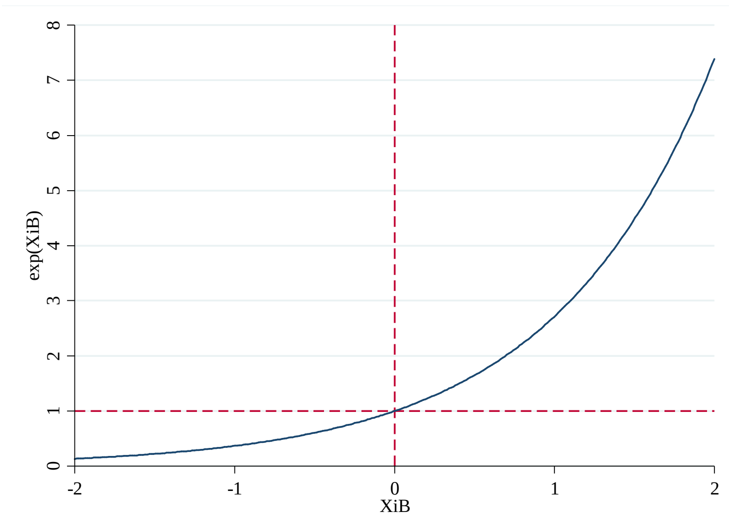
$\mathbf{x}_i\boldsymbol{\beta} \rightarrow \infty$?

$\mathbf{x}_i\boldsymbol{\beta} = 0$?

$\mathbf{x}_i\boldsymbol{\beta} \rightarrow -\infty$?

$$Note: e^{-\infty} = \frac{1}{e^{\infty}}$$

e^{XiB}



Probability P as a function of X

$$P_i = \left(\frac{e^{\mathbf{x}_i\boldsymbol{\beta}}}{1+e^{\mathbf{x}_i\boldsymbol{\beta}}}\right)\left(\frac{1/e^{\mathbf{x}_i\boldsymbol{\beta}}}{1/e^{\mathbf{x}_i\boldsymbol{\beta}}}\right) = \frac{1}{\frac{1}{e^{\mathbf{x}_i\boldsymbol{\beta}}}+1} = \frac{1}{1+e^{-\mathbf{x}_i\boldsymbol{\beta}}}$$

$$1-P_i = 1-\frac{e^{\mathbf{x}_i\boldsymbol{\beta}}}{1+e^{\mathbf{x}_i\boldsymbol{\beta}}} = \frac{1}{1+e^{\mathbf{x}_i\boldsymbol{\beta}}}$$

- View "Logit function of XiB.xls"
- β_1 shifts curve left or right
- β_2 shifts slope at all points
- But how will we estimate the parameters?

Use MLE

$$\text{Sort } Y_i \text{ so that } Y_i = \begin{cases} 1 \text{ in observatios } 1, ..., 200 \\ 0 \text{ in observations } 201, ..., 300 \end{cases}$$

$$\mathcal{L} = (P_1)(P_2)\cdots(P_{200})(1-P_{201})\cdots(1-P_{300})$$

$$\mathcal{L} = \left[\prod_{i=1}^{200}(P_i)\right]\left[\prod_{i=201}^{300}(1-P_i)\right]$$

What is being assumed here?

$$\mathcal{L} = \prod_{i=1}^{300} P_i^{Y_i}(1-P_i)^{(1-Y_i)}$$

Now it doesn't matter how Y is sorted

The Log of the Likelihood

$$\mathcal{L} = \prod_{i=1}^{N} P_i^{Y_i} (1 - P_i)^{(1-Y_i)} = \prod_{Y_i=1} (P_i) \prod_{Y_i=0} (1 - P_i)$$

$$\ln \mathcal{L} = \sum_{i=1}^{N} Y_i \ln P_i + \sum_{i=1}^{N} (1 - Y_i) \ln (1 - P_i)$$

$$= \sum_{Y_i=0} \ln P_i + \sum_{Y_i=1} \ln (1 - P_i)$$

$$\ln \mathcal{L} = \sum_{Y_i=1} \ln \left(\frac{1}{1 + e^{-\mathbf{x}_i \boldsymbol{\beta}}} \right) + \sum_{Y_i=0} \ln \left(\frac{1}{1 + e^{\mathbf{x}_i \boldsymbol{\beta}}} \right)$$

Maximize the Log Likelihood

$$\ln L = \sum_{Y_i=1} \ln \left(\frac{1}{1 + e^{-\mathbf{x}_i \hat{\boldsymbol{\beta}}}} \right) + \sum_{Y_i=0} \ln \left(\frac{1}{1 + e^{\mathbf{x}_i \hat{\boldsymbol{\beta}}}} \right)$$

Maximize $\ln L$ with respect to $\hat{\beta}_1, \hat{\beta}_2, \ldots$

No analytic solution, so the answer must be found using a maximization algorithm (Long 3.6)

- The MLE estimates
  - are consistent
  - asymptotically efficient
  - asymptotically normal
- Taken together, this means you can do confidence intervals and hypothesis tests on the coefficients.

How big does N have to be?

- Nobody really knows
- Long:
  - N > 500 is probably good enough, but it depends on the data (collinearity, etc.)
  - 100 < N < 500 marginal
  - N < 100 risky, properties basically unknown
- If convergence takes many tries, the lnL function is basically flat or lumpy.
- Example

Marginal Effects: LPM

$$\hat{P}_i = \hat{\beta}_1 + \hat{\beta}_2 age_i + u_i \qquad \frac{d\hat{P}_i}{d\left(age_i\right)} = \hat{\beta}_2$$

$$\hat{P}_i = \mathbf{x}_i\hat{\boldsymbol{\beta}} = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \ldots + \hat{\beta}_k X_{ki} \qquad \frac{d\hat{P}_i}{dX_{ji}} = \hat{\beta}_j$$

$$\begin{aligned}\hat{P}_{10} &= -0.26 + 0.063 age_i \\ &= -0.26 + 0.063(10) \\ &= -0.26 + 0.63 \\ &= 0.37\end{aligned} \qquad \begin{aligned}\hat{P}_{11} &= -0.26 + 0.063 age_i \\ &= -0.26 + 0.063(11) \\ &= -0.26 + 0.693 \\ &= 0.433\end{aligned}$$

$$\hat{P}_{11} - \hat{P}_{10} = 0.063$$

Marginal Effects: Logit

$$\hat{P}_i = \frac{1}{1+e^{-\left(\hat{\beta}_1+\hat{\beta}_2 X_i\right)}}$$

$$\frac{d\hat{P}_i}{dX_i} = \hat{P}_i\left(1-\hat{P}_i\right)\hat{\beta}_2$$

$$Example: Y_i = -22 + 1.8\left(age_i\right)$$

$$\frac{d\hat{P}_i}{d\left(age_i\right)} = \hat{P}_i\left(1-\hat{P}_i\right)(1.8)$$

$$\mathbf{x}_i\hat{\boldsymbol{\beta}} = -22 + 1.8(10) = -4$$

$$\hat{P}_i = \frac{1}{1+e^{-(-4)}} = \frac{1}{1+e^{4}} = 0.018$$

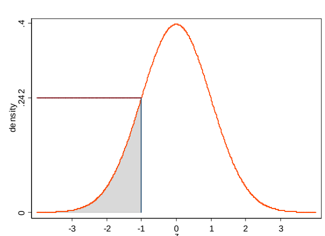
$$\frac{d\hat{P}_i}{d\left(age_i\right)} = (0.018)(0.982)1.8 = 0.032$$

General case: $\dfrac{\partial P_i}{\partial X_{ji}} = P_i\left(1-P_i\right)\beta_j$

age_i	$\hat{P}$	$\frac{d\hat{P}_i}{d\left(age_i\right)}$
10	0.018	0.032
11		
12		
13		
14		

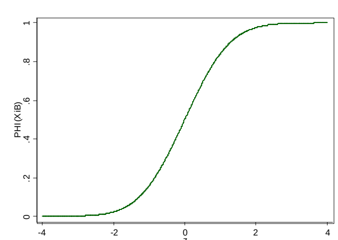
Alternative Function: Probit

Recall: $$\mathcal{L} = \prod_{i=1}^{N} P_i^{Y_i}\left(1-P_i\right)^{(1-Y_i)} = \prod_{Y_i=1}\left(P_i\right)\prod_{Y_i=0}\left(1-P_i\right)$$



Normal Density

$\phi(z)$



Cumulative Normal

$\Phi(z)$

Cumulative normal looks similar to the logit curve, effect is non-linear and diminishes as P approaches zero or 1.

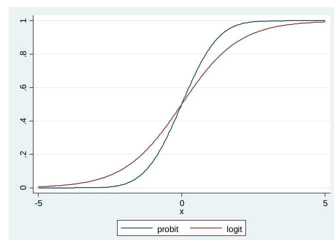
Probit Regression

Let $P_i = \Phi(\mathbf{x}_i\boldsymbol{\beta}) = \Phi(\beta_1 + \beta_2 X_i)$

$$L = \prod_{Y_i=1}\left[\Phi\left(\mathbf{x}_i\hat{\boldsymbol{\beta}}\right)\right]\prod_{Y_i=0}\left[1-\Phi\left(\mathbf{x}_i\hat{\boldsymbol{\beta}}\right)\right]$$

$$\ln L = \sum_{Y_i=1}\left[\ln\Phi\left(\mathbf{x}_i\hat{\boldsymbol{\beta}}\right)\right]+\sum_{Y_i=0}\left[\ln\left(1-\Phi\left(\mathbf{x}_i\hat{\boldsymbol{\beta}}\right)\right)\right]$$

Maximize lnL with respect to $\hat{\boldsymbol{\beta}}$



Compare Probit and Logit

Marginal Effects: Probit

$$\hat{P}_i = \Phi\left(\hat{\beta}_1 + \hat{\beta}_2 X_i\right) \quad \frac{d\hat{P}_i}{dX_i} = \phi\left(\hat{\beta}_1 + \hat{\beta}_2 X_i\right)\hat{\beta}_2$$

$$\mathbf{x}_i\boldsymbol{\beta} = -12 + 1(age_i) = -12 + 10 = -2$$

$$\hat{P}_{10} = \Phi(-2) = 0.023$$

$$\frac{dP_i}{dX_i} = \phi(-2)(1) = 0.054$$

$$\frac{\partial \hat{P}_i}{\partial X_{ji}} = \phi\left(\mathbf{x}_i\hat{\boldsymbol{\beta}}\right)\hat{\beta}_j = \phi\left(\hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \ldots + \hat{\beta}_k X_{ki}\right)\hat{\beta}_j$$