

**PPOL 503-03, PPOL 503-04, Fall 2016**  
**Course Notes #6: Binary Probit**

**I. Binary Probit Model Specification**

$$I = \beta X$$

$$I^* \sim N(0, 1)$$

$$Y = 1 \text{ if } I_t > I_t^* \\ = 0 \text{ if } I_t \leq I_t^*$$

The probit model with a single regressor is:

$$\Pr(Y = 1 | X) = \Phi(\beta_0 + \beta_1 X)$$

where  $\Phi$  is the cumulative standard normal distribution function.

Suppose for example  $Y = 1$  if the mortgage is denied;  $X$  is the payment to income ratio ( $P|I$ ),  $\beta_0 = -2$  and  $\beta_1 = 3$ . What is the probability of denial if the ( $P|I$ ) = 0.4? This probability is  $\Phi(-2 + 3(P|I)) = \Phi(-2 + 3(0.4)) = \Phi(-0.8) = \Pr(Z \leq -0.8) = 21.2\%$ . That is when the  $P|I$  is 0.4, the predicted probability that the application will be denied is 21.2%.

The probit coefficient  $\beta_1$  is the change in the  $z$ -value associated with a unit change in  $X$ . If  $\beta_1$  is positive, an increase in  $X$  increases the  $z$ -value and thus decreases the probability that  $Y = 1$ . If  $\beta_1$  is negative, an increase in  $X$  decreases the probability that  $Y = 1$ . Although the effect of  $X$  on the  $z$ -value is linear, its effect on the probability is non-linear.

The estimated probit function has a stretched “S” shape. It is nearly zero and flat for small values of the  $P|I$  ratio, it turns and increases for intermediate values, and it flattens out and is nearly one for large values. For example, if the  $P|I$  is 0.2, the probability of denial is 2.1%. For  $P|I = 0.6$ , the probability of denial is 98.3%.

**II. Probit Regression with Multiple Regressors**

$$\Pr(Y = 1 | X_1, X_2, X_3, X_4) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4)$$

The predicted probability that  $Y = 1$ , given the values of  $X_1, X_2, X_3, X_4$  is calculated by computing the  $z$ -value,  $z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$  and then looking up this  $z$ -value in the normal distribution table.

The coefficient  $\beta_1$  is the change in the  $z$ -value arising from a unit change in  $X_1$  holding constant  $X_2, X_3, X_4$ . The effect on the predicted probability of a change in a regressor is computed by (1) computing the predicted probability for the initial value of the regressors, (2) computing the predicted probability for the new or changed value of the regressors and (3) taking their difference.

Example: What is the effect of race on the probability of mortgage denial holding constant the payment-to-income ratio? To answer this question, one estimates a probit model predicting the probability of denial with both the *P/I ratio* and *black* as regressors:

$$\Pr(\text{deny} = 1) = \Phi(-2.26 + 2.74 \text{ P/I ratio} + 0.71 \text{ black})$$

(0.16)    (0.44)            (0.083)

The values of the coefficients are difficult to interpret but the sign and statistical significance are not. The coefficient on *black* is highly significant, indicating that an African American applicant has a higher probability of denial than a white applicant, holding constant the payment to income ratio. (The  $t$ -statistic for the *black* coefficient is 8.55.) For a white applicant with *P/I ratio* = 0.3, the predicted probability of denial is 7.5%, while for a black applicant with a *P/I ratio* = 0.3, it is 23.3%; the difference in denial probabilities between these two hypothetical applicants is 15.8 percentage points.

### III. Another Probit Regression Example

The data in this example were gathered on undergraduates applying to graduate school and includes undergraduate GPAs, the reputation of the school of the undergraduate (a topnotch indicator), the students' GRE score, and whether or not the student was admitted to graduate school. Using this dataset, we can predict admission to graduate school using undergraduate GPA, GRE scores, and the reputation of the school of the undergraduate. Our outcome variable is binary, and we will use a probit model. Thus, our model will calculate a predicted probability of admission based on our predictors. The probit model does so using the cumulative distribution function of the standard normal.

```

summarize
  Variable |      Obs      Mean   Std. Dev.   Min   Max
-----+-----
    admit |     400     .3175   .4660867     0     1
      gre |     400    587.7   115.5165    220    800
  topnotch |     400     .1625   .3693709     0     1
      gpa |     400    3.3899   .3805668    2.26     4
tabulate admit
  admit |      Freq.      Percent      Cum.
-----+-----
      0 |         273         68.25         68.25
      1 |         127         31.75        100.00
-----+-----
  Total |         400        100.00

```

To run the model in Stata, we first give the response variable (**admit**), followed by our predictors (**gre**, **topnotch** and **gpa**).

```

probit admit gre topnotch gpa

```

```

Iteration 0:  log likelihood = -249.98826
Iteration 1:  log likelihood = -238.97735
Iteration 2:  log likelihood = -238.94339
Iteration 3:  log likelihood = -238.94339

```

```

Probit regression                               Number of obs   =    400
                                                LR chi2(3)      =   22.09
                                                Prob > chi2     =   0.0001
Log likelihood = -238.94339                    Pseudo R2      =   0.0442

```

```

-----+-----
  admit |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
      gre |   .0015244   .0006382     2.39   0.017   .0002736   .0027752
  topnotch |   .2730334   .1795984     1.52   0.128   -.078973   .6250398
      gpa |   .4009853   .1931077     2.08   0.038   .0225012   .7794694
      cons |  -2.797884   .6475363    -4.32   0.000   -4.067032  -1.528736
-----+-----

```

## Iteration History

```

Iteration 0:  log likelihood = -249.98826
Iteration 1:  log likelihood = -238.97735
Iteration 2:  log likelihood = -238.94339
Iteration 3:  log likelihood = -238.94339

```

**Iteration History** - This is a listing of the log likelihoods at each iteration for the probit model. Remember that probit regression uses maximum likelihood

estimation, which is an iterative procedure. The first iteration (called Iteration 0) is the log likelihood of the "null" or "empty" model; that is, a model with no predictors. At the next iteration (called Iteration 1), the specified predictors are included in the model. In this example, the predictors are **gre**, **topnotch** and **gpa**. At each iteration, the log likelihood increases because the goal is to maximize the log likelihood. When the difference between successive iterations is very small, the model is said to have "converged" and the iterating stops.

**LR chi2(3)** - This is the Likelihood Ratio (LR) Chi-Square test that at least one of the predictors' regression coefficient is not equal to zero. The number in the parentheses indicates the degrees of freedom of the Chi-Square distribution used to test the LR Chi-Square statistic and is defined by the number of predictors in the model (3).

**Std. Err.** - These are the standard errors of the individual regression coefficients. They are used in both the calculation of the **z** test statistic, superscript *j*, and the confidence interval of the regression coefficient, superscript *l*.

**z** - The test statistic **z** is the ratio of the **Coef.** to the **Std. Err.** of the respective predictor. The **z** value follows a standard normal distribution which is used to test against a two-sided alternative hypothesis that the **Coef.** is not equal to zero.

**P>|z|** - This is the probability the **z** test statistic (or a more extreme test statistic) would be observed under the null hypothesis that a particular predictor's regression coefficient is zero, given that the rest of the predictors are in the model. For a given alpha level, **P>|z|** determines whether or not the null hypothesis can be rejected. If **P>|z|** is less than alpha, then the null hypothesis can be rejected and the parameter estimate is considered statistically significant at that alpha level.

**gre** - The **z** test statistic for the predictor **gre** is  $(0.0015244/0.0006382) = 2.39$  with an associated p-value of 0.017. If we set our alpha level to 0.05, we would reject the null hypothesis and conclude that the regression coefficient for **gre** has been found to be statistically different from zero given **topnotch** and **gpa** are in the model.

**topnotch** - The **z** test statistic for the predictor **topnotch** is  $(0.2730334/0.1795984) = 1.52$  with an associated p-value of 0.128. If we set our alpha level to 0.05, we would fail to reject the null hypothesis and conclude that the regression coefficient for **topnotch** has not been found to be statistically different from zero given **gre** and **gpa** are in the model.

**gpa** - The **z** test statistic for the predictor **gpa** is  $(0.4009853/0.1931077) = 2.08$  with an associated p-value of 0.038. If we set our alpha level to 0.05, we would reject the null hypothesis and conclude that the regression coefficient for **gpa** has been found to be statistically different from zero given **gre** and **topnotch** are in the model.

**cons** -The **z** test statistic for the intercept, **\_cons**, is  $(-2.797884/0.6475363) = -4.32$  with an associated p-value of  $< 0.001$ . With an alpha level of 0.05, we would reject the null hypothesis and conclude that **\_cons** has been found to be statistically different from zero given **gre**, **topnotch** and **gpa** are in the model and evaluated at zero.

The predicted probability of admission can be calculated using these coefficients. For a given record, the predicted probability of admission is

$$F(-2.797884 + gre \times 0.0015244 + topnotch \times 0.2730334 + gpa \times 0.4009853)$$

where  $F$  is the cumulative distribution function of the standard normal. However, interpretation of the coefficients in probit regression is not as straightforward as the interpretations of coefficients in linear regression or logit regression. The increase in probability attributed to a one-unit increase in a given predictor is dependent both on the values of the other predictors and the starting value of the given predictors. For example, if we hold **gre** and **topnotch** constant at zero, the one unit increase in **gpa** from 2 to 3 has a different effect than the one unit increase from 3 to 4 (note that the probabilities do not change by a common difference or common factor):

$$F(-2.797884 + 2 \times 0.4009853) = 0.02297167$$

$$F(-2.797884 + 3 \times 0.4009853) = 0.05536416$$

$$F(-2.797884 + 4 \times 0.4009853) = 0.11625017$$

and the effects of these one unit increases are different if we hold **gre** and **topnotch** constant at their respective means instead of zero:

$$F(-2.797884 + 587.7 \times 0.0015244 + 0.1625 \times 0.2730334 + 2 \times 0.4009853) = 0.1455628$$

$$F(-2.797884 + 587.7 \times 0.0015244 + 0.1625 \times 0.2730334 + 3 \times 0.4009853) = 0.25634003$$

$$F(-2.797884 + 587.7 \times 0.0015244 + 0.1625 \times 0.2730334 + 4 \times 0.4009853) = 0.39986946.$$

To generate values from  $F$  in Stata, use the **normal** function. For example,

```
display normal(0)
```

will display .5, indicating that  $F(0) = .5$  (i.e., half of the area under the standard normal distribution curve falls to the left of zero). The first student in our dataset has a GRE score of 380, a GPA of 3.61, and a topnotch indicator value of 0. We could multiply these values by their corresponding coefficients,

```
display -2.797884 + (.0015244*380) + (.2730334*0) +  
(.4009853*3.61)
```

to determine that the predicted probability of admittance is  $F(-0.77105507)$ . To find this value, we type

```
display normal(-0.77105507)
```

and arrive at a predicted probability of 0.22033715.

#### **IV. Logit Versus Probit: Which to Use?**

1. The logistic distribution is very similar to the normal except in the extremes (that is, with very small or very large values.) For intermediate values, the two distributions yield similar probabilities.
2. Compared to the normal distribution, the logistic assigns larger probabilities to  $Y = 0$  when  $\beta_i X_i$  is small and smaller probabilities to  $Y = 1$  when  $\beta_i X_i$  is large.
3. For example, if the value of the cumulative probability function has a Z score of -3, the probability assigned under the cumulative logistic function is .0474 while under the cumulative normal the probability assigned to a Z score of -3 is .0013. On the other hand, if the value of the cumulative probability function is 3, the probability assigned under the cumulative normal is .998 while under the cumulative logistic the value is .953.
4. Although the two distributions yield similar results, the parameter estimates ( $\beta_i$ ) are not directly comparable. Rule of thumb (Amemiya): the logit coefficient multiplied by .625 is a good approximation of the probit coefficient. Alternatively, one can divide the probit estimates by 1.6.

5. The logit coefficients should be divided by 4 to make them roughly comparable to the LPM estimates. The probit estimates should be divided by 2.5 to make them comparable to the LPM estimates.

EXAMPLE: Married Women’s Labor Force Participation: Comparison of LPM, Logit and Probit

Dependent Variable: LFP = 1 if in the labor force; = 0 if not

Independent Variables	LPM (OLS)	LOGIT (MLE)	PROBIT (MLE)
Nwifeincome	-.0034*** (.0015)	-.021*** (.008)	-.012*** (.005)
Education	.038*** (.007)	.221*** (.043)	.131*** (.025)
Experience	.039*** (.006)	.206*** (.032)	.123*** (.019)
Experience <sup>2</sup>	-.0006*** (.00018)	-.0032** (.0010)	-.0019*** (.0006)
AGE	-.016*** (.002)	-.088*** (.015)	-.053*** (.008)
KIDLT6	-.262*** (.032)	-1.443*** (.204)	-.868*** (.119)
KIDGE6	.013 (.013)	.060 (.075)	.036 (.043)
% Correctly Classified	73.4%	73.6%	73.4%
Pseudo R <sup>2</sup>	.264	.220	.221

1. The results reported in Table A compare the LPM. Logit and probit estimation for married women’s labor force participation.

2. The three models tell a consistent story; the signs of the coefficients are the same across the models and the same variables are statistically significant.

3. The magnitudes of the effects are not directly comparable. The rule of thumb is to divide the logit estimates by 4 and the probit estimates by 2.5 to make them comparable to the LPM estimates.

4. KIDSLT6: scaled logit (-.361); scaled probit (-.347); both are larger in magnitude than the LPM coefficient of (-.261)
5. EDUCATION: scaled logit (.055); scaled probit (.052); both are somewhat larger than the LPM estimate of .038.
6. The biggest difference between the LPM model and the logit and probit models is that the LPM assumes constant marginal effects for KIDSLT6, EDUCATION etc, whereas the logit and probit models imply diminishing magnitudes of the partial effects.
7. For the LPM model, one more small child is estimated to reduce the probability of LF participation by about .262 (26.2% points) regardless of how many young children the woman already has.
8. Consider the marginal effects from the probit model for a woman with the following characteristics:  
Nwifeincome = 20.13, education = 12.3, experience = 10.6, and age = 42.5 (these are the sample averages). Question: what is the decrease in the probability of working in going from zero to one child?
9. Evaluate the CDF  $\Phi(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)$  with KIDSLT6 = 0 and KIDSLT6 = 1.

All other variables are set to the sample means.

Result:  $.373 - .707 = -.334$

This means that the probability of labor force participation is .334 or 33.4% points lower when a woman has one child under 6 compared to no child under age 6.

## V. Testing Restrictions for Logit and Probit Models

- F-tests do not apply for these models. Instead, we use the *likelihood ratio statistic*, which closely corresponds to the F-statistic we used with OLS.
- The *LR* test compares the log-likelihood function of the unrestricted model (the model with all the variables included) to the log-likelihood function of the restricted model (the model with all – or a subset of – the variables

assumed to have no effect on the dependent variable). Because MLE maximizes the log-likelihood function, dropping variables generally leads to a smaller log-likelihood. The question is whether the fall in the log-likelihood is large enough to conclude that the dropped variables are important. The LR test is a way of testing whether the difference is significant (actually, test is whether  $2(\text{Loglikelihood}_{UR} - \text{Loglikelihood}_R)$  is significant, using a chi-square distribution).

- In Stata, as you did with OLS, you can use the “test” command after the logit or probit command to test different restrictions.

**EXAMPLE:** the decision to purchase a foreign or American made car;  $Y = 1$  if you purchase a foreign car and 0 if you opt for an American manufactured car. Explanatory variables include weight, length and MPG.

You want to test two restrictions: 1) that the effects of weight and length are jointly equal to zero; and 2) the effect of weight equals the effect of length.

```
. probit foreign weight length mpg
```

```
Iteration 0:  log likelihood = -45.03321
Iteration 1:  log likelihood = -29.202588
Iteration 2:  log likelihood = -27.046012
Iteration 3:  log likelihood = -26.845695
Iteration 4:  log likelihood = -26.843111
Iteration 5:  log likelihood = -26.843111
```

```
Probit estimates                               Number of obs   =          74
                                                LR chi2(3)      =          36.38
                                                Prob > chi2     =          0.0000
Log likelihood = -26.843111                    Pseudo R2      =          0.4039
```

foreign	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
weight	-.002292	.0010912	-2.10	0.036	-.0044308 - .0001533
length	-.0015517	.0333588	-0.05	0.963	-.0669337 .0638303
mpg	-.1040006	.0515543	-2.02	0.044	-.2050451 -.0029561
_cons	8.437624	4.327984	1.95	0.051	-.0450683 16.92032

```
. test weight length
```

```
( 1)  weight = 0.0
( 2)  length = 0.0
```

```
      chi2( 2) =    17.03
Prob > chi2 =    0.0002
```

```
. test weight=length

( 1)  weight - length = 0.0

      chi2( 1) =    0.00
      Prob > chi2 =  0.9828
```

You reject the null hypothesis that weight and length are jointly equal to zero, but you fail to reject the null hypothesis that the effect of weight equals the effect of length.

## VI. Testing for Structural Breaks (Slope Effects) in Logit and Probit Models

- Sometimes a particular variable may have slope effects, that is, there may be structural breaks in your data. In this case it may be appropriate to stratify the data by a specific characteristic (i.e. gender or race). Pooling the data imposes the restriction that an independent variable has the same effect on the dependent variable when the opposite may be true.
- You can conduct a likelihood ratio test:

$$LR [ i + 1 ] = [ 2 \loglikelihood (sample 1) + 2 \loglikelihood (sample 2) ] - 2 \loglikelihood (pooled model)$$

$$LR [ i + 1 ] = [ 2 \loglikelihood (sample 1) + 2 \loglikelihood (sample 2) ] - 2 \loglikelihood (pooled model)$$

Where samples 1 and 2 are pooled and i is the number of independent variables.

EXAMPLE: Is evacuation behavior from Hurricanes Dennis and Floyd statistically equivalent?

Logit Coefficient

Variable	Floyd Model	Dennis Model	Pooled Model
PETS	-0.66	-1.20	-0.79
MOBILEHOME	1.56	2.00	1.62
TENURE	-0.02	-0.02	-0.02
EDUC	0.05	-0.04	0.02
Constant	-0.92	-0.78	-0.97

2 Log Likelihood (Constant only)	-687.36	-440.87	-1186.64
2 Log Likelihood (With Covariates)	-641.84	-382.84	-1095.26
Model Chi-Square	45.52	58.03	91.38

Chi-Square [(Floyd + Dennis) - Pooled] =

$$(-641.84 + (-382.84)) - (-1095.26) = 71.58 \quad DF = 5$$

Chi-Square Critical Value = 13.277 (p = .01)

- Since the Chi-square value is greater than the critical value the set of coefficients are statistically different. The pooled model is inappropriate.