Getting Started in Logit and Ordered Logit Regression
(v. 3.1)

Oscar Torres-Reyna
otorres@princeton.edu
Logit model

• Use logit models whenever your dependent variable is binary (also called dummy) which takes values 0 or 1.

• Logit regression is a nonlinear regression model that forces the output (predicted values) to be either 0 or 1.

• Logit models estimate the probability of your dependent variable to be 1 ($Y=1$). This is the probability that some event happens.
Logit model

From Stock & Watson, key concept 9.3. The logit model is:

\[
\Pr(Y = 1 \mid X_1, X_2, \ldots, X_k) = F(\beta_0 + \beta_1X_1 + \beta_2X_2 + \ldots + \beta_KX_K)
\]

\[
\Pr(Y = 1 \mid X_1, X_2, \ldots, X_k) = \frac{1}{1 + e^{- (\beta_0 + \beta_1X_1 + \beta_2X_2 + \ldots + \beta_KX_K)}}
\]

\[
\Pr(Y = 1 \mid X_1, X_2, \ldots, X_k) = \frac{1}{1 + \left(\frac{1}{e^{(\beta_0 + \beta_1X_1 + \beta_2X_2 + \ldots + \beta_KX_K)}}\right)}
\]

Logit and probit models are basically the same, the difference is in the distribution:

- Logit – Cumulative standard logistic distribution \((F)\)
- Probit – Cumulative standard normal distribution \((\Phi)\)

Both models provide similar results.
In Stata you run the model as follows:

```
.logit y_bin x1 x2 x3 x4 x5 x6 x7
```

It tests whether the combined effect, of all the variables in the model, is different from zero. If, for example, $p < 0.05$ then the model have some relevant explanatory power, which does not mean it is well specified or at all correct.

### Logistic regression

- **Number of obs**: 490
- **LR chi2(7)**: 183.06
- **Prob > chi2**: 0.0000
- **Pseudo R2**: 0.3633

| y_bin | Coef. | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|-------|-------|-----------|------|-----|---------------------|
| x1    | 0.2697623 | 0.1759677 | 1.53 | 0.125 | -0.0751281 to 0.6146527 |
| x2    | -0.2500592 | 0.1459846 | -1.71 | 0.087 | -0.5361837 to 0.030653 |
| x3    | 0.1150445 | 0.1486181 | 0.77 | 0.439 | -0.1762417 to 0.4063306 |
| x4    | 0.3649722 | 0.153434 | 2.38 | 0.017 | 0.0642472 to 0.6656973 |
| x5    | -0.3131214 | 0.1467796 | -2.13 | 0.033 | -0.6008042 to -0.0254386 |
| x6    | -0.1361499 | 0.1566993 | -0.87 | 0.385 | -0.4432749 to 0.1709752 |
| x7    | 3.206987 | 0.3631481 | 8.83 | 0.000 | 2.495229 to 3.918744 |
| _cons | 1.58614 | 0.39927 | 3.97 | 0.000 | 0.803585 to 2.368695 |

Note: 1 failure and 1 success completely determined.

Logit coefficients are in log-odds units and cannot be read as regular OLS coefficients. To interpret you need to estimate the predicted probabilities of $Y=1$ (see next page).

Test the hypothesis that each coefficient is different from 0. To reject this, the t-value has to be higher than 1.96 (for a 95% confidence). If this is the case then you can say that the variable has a significant influence on your dependent variable ($y$). The higher the $z$ the higher the relevance of the variable.

Two-tail p-values test the hypothesis that each coefficient is different from 0. To reject this, the p-value has to be lower than 0.05 (95%, you could choose also an alpha of 0.10), if this is the case then you can say that the variable has a significant influence on your dependent variable ($y$).
After running the model:

`logit y_bin x1 x2 x3 x4 x5 x6 x7`

Type

`predict y_bin_hat` /*These are the predicted probabilities of Y=1 */

Here are the estimations for the first five cases, type:

`browse y_bin x1 x2 x3 x4 x5 x6 x7 y_bin_hat`

<table>
<thead>
<tr>
<th>y_bin</th>
<th>x1</th>
<th>x2</th>
<th>x3</th>
<th>x4</th>
<th>x5</th>
<th>x6</th>
<th>x7</th>
<th>y_bin_hat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>.2779036</td>
<td>-1.107956</td>
<td>.2825536</td>
<td>-2.971267</td>
<td>.554832</td>
<td>-5.820704</td>
<td>.7841014</td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>.3205847</td>
<td>-9.94872</td>
<td>.4925385</td>
<td>-1.371243</td>
<td>-.0959273</td>
<td>-9.641485</td>
<td>.6673266</td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>.3634657</td>
<td>-.789484</td>
<td>.7025234</td>
<td>.2287798</td>
<td>-.7466869</td>
<td>-.7462227</td>
<td>.5267279</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>.245144</td>
<td>-.885533</td>
<td>-.0943909</td>
<td>-.3198499</td>
<td>-.3573879</td>
<td>.0628607</td>
<td>.9274359</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>.424623</td>
<td>-.7297683</td>
<td>.9461306</td>
<td>.1230506</td>
<td>-.0358964</td>
<td>.095743</td>
<td>.9439594</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>.4772141</td>
<td>-.723246</td>
<td>1.02568</td>
<td>.1175985</td>
<td>-.0022627</td>
<td>.0965806</td>
<td>.9448991</td>
</tr>
</tbody>
</table>

To estimate the probability of Y=1 for the first row, replace the values of X into the logit regression equation. For the first case, given the values of X there is 79% probability that Y=1:

\[
Pr(Y = 1 | X_1, X_2, \ldots, X_7) = \frac{1}{1 + e^{-(1.58 + 0.26X_1 - 0.25X_2 + 0.11X_3 + 0.36X_4 - 0.31X_5 - 0.13X_6 + 3.20X_7)}} = 0.7841014
\]
It tests whether the combined effect, of all the variables in the model, is different from zero. If, for example, < 0.05 then the model have some relevant explanatory power, which does not mean it is well specified or at all correct.

Logistic regression

Log likelihood = -160.442

|     | Odds Ratio | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|-----|------------|-----------|-------|-----|----------------------|
| x1  | 1.309653   | .2304567  | 1.53  | 0.125 | .9276246, 1.849014   |
| x2  | .7787547   | .1136862  | -1.71 | 0.087 | .5849765, 1.036724   |
| x3  | 1.121923   | .1667381  | 0.77  | 0.439 | .8384153, 1.501299   |
| x4  | 1.440474   | .2210176  | 2.38  | 0.017 | 1.066356, 1.945847   |
| x5  | .731162    | .1073196  | -2.13 | 0.033 | .5483705, .9748823   |
| x6  | .8727118   | .1367534  | -0.87 | 0.385 | .6419307, 1.186461   |
| x7  | 24.70453   | 8.971405  | 8.83  | 0.000 | 12.12451, 50.33718   |

Note: 1 failure and 1 success completely determined.

They represent the odds of Y=1 when X increases by 1 unit. These are the exp(logit coeff).
If the OR > 1 then the odds of Y=1 increases
If the OR < 1 then the odds of Y=1 decreases
Look at the sign of the logit coefficients

Test the hypothesis that each coefficient is different from 1. To reject this, the t-value has to be higher than 1.96 (for a 95% confidence). If this is the case then you can say that the variable has a significant influence on your dependent variable (y). The higher the z the higher the relevance of the variable.

Two-tail p-values test the hypothesis that each coefficient is different from 1. To reject this, the p-value has to be lower than 0.05 (95%). you could choose also an alpha of 0.10, if this is the case then you can say that the variable has a significant influence on your dependent variable (y).
Predicted probabilities and marginal effects

For the latest procedure see the following document:

http://dss.princeton.edu/training/Margins.pdf

The procedure using `prvalue` in the following pages does not work with Stata 13.
Ordinal logit

When a dependent variable has more than two categories and the values of each category have a meaningful sequential order where a value is indeed ‘higher’ than the previous one, then you can use ordinal logit.

Here is an example of the type of variable:

<table>
<thead>
<tr>
<th>Agreement Level</th>
<th>Freq.</th>
<th>Percent</th>
<th>Cum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disagree</td>
<td>190</td>
<td>38.78</td>
<td>38.78</td>
</tr>
<tr>
<td>Neutral</td>
<td>104</td>
<td>21.22</td>
<td>60.00</td>
</tr>
<tr>
<td>Agree</td>
<td>196</td>
<td>40.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

| Total           | 490   | 100.00  |     |
Ordinal logit: the setup

Ordered logistic regression

```
. ologit y_ordinal x1 x2 x3 x4 x5 x6 x7
```

Log likelihood = -458.38145

Dependent variable: y_ordinal
Independent variable(s): x1 x2 x3 x4 x5 x6 x7

Logit coefficients are in log-odds units and cannot be read as regular OLS coefficients. To interpret you need to estimate the predicted probabilities of Y=1 (see next page).

Ancillary parameters to define the changes among categories (see next page)

Test the hypothesis that each coefficient is different from 0. To reject this, the t-value has to be higher than 1.96 (for a 95% confidence). If this is the case then you can say that the variable has a significant influence on your dependent variable (y). The higher the z the higher the relevance of the variable.

Two-tail p-values test the hypothesis that each coefficient is different from 0. To reject this, the p-value has to be lower than 0.05 (95%, you could choose also an alpha of 0.10), if this is the case then you can say that the variable has a significant influence on your dependent variable (y).

If this number is < 0.05 then your model is ok. This is a test to see whether all the coefficients in the model are different than zero.
Ordinal logit: predicted probabilities

Following Hamilton, 2006, p.279, ologit estimates a score, S, as a linear function of the X’s:

\[ S = 0.22X_1 - 0.05X_2 + 0.11X_3 + 0.22X_4 - 0.29X_5 + 0.003X_6 + 1.57X_7 \]

Predicted probabilities are estimated as:

\[
\begin{align*}
P(\text{y_ordinal=“disagree”}) &= P(S + u \leq _\text{cut1}) = P(S + u \leq -0.5528058) \\
P(\text{y_ordinal=“neutral”}) &= P(_\text{cut1} < S + u \leq _\text{cut2}) = P(-0.5528058 < S + u \leq 0.5389237) \\
P(\text{y_ordinal=“agree”}) &= P(_\text{cut2} < S + u) = P(0.5389237 < S + u)
\end{align*}
\]

To estimate predicted probabilities type `predict` right after `ologit` model. Unlike `logit`, this time you need to specify the predictions for all categories in the ordinal variable (y_ordinal), type:

`predict disagree neutral agree`
Ordinal logit: predicted probabilities

To read these probabilities, as an example, type

```
browse country disagree neutral agree if year==1999
```

In 1999 there is a 62% probability of ‘agreement’ in Australia compared to 58% probability in ‘disagreement’ in Brazil while Denmark seems to be quite undecided.

<table>
<thead>
<tr>
<th>country</th>
<th>disagree</th>
<th>neutral</th>
<th>agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>.1700809</td>
<td>.2090298</td>
<td>.6208892</td>
</tr>
<tr>
<td>Austria</td>
<td>.17576</td>
<td>.2127421</td>
<td>.6114978</td>
</tr>
<tr>
<td>Belgium</td>
<td>.3058564</td>
<td>.2617683</td>
<td>.4323753</td>
</tr>
<tr>
<td>Botswana</td>
<td>.1215602</td>
<td>.1703741</td>
<td>.7080657</td>
</tr>
<tr>
<td>Brazil</td>
<td>.5808533</td>
<td>.2241725</td>
<td>.1949743</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>.3134856</td>
<td>.2628762</td>
<td>.4236383</td>
</tr>
<tr>
<td>Burundi</td>
<td>.5940011</td>
<td>.2193996</td>
<td>.1865993</td>
</tr>
<tr>
<td>Canada</td>
<td>.1627286</td>
<td>.2039865</td>
<td>.6332849</td>
</tr>
<tr>
<td>Chile</td>
<td>.1998139</td>
<td>.2267881</td>
<td>.5733979</td>
</tr>
<tr>
<td>Denmark</td>
<td>.3604209</td>
<td>.2663039</td>
<td>.3732751</td>
</tr>
</tbody>
</table>
Predicted probabilities and marginal effects

For the latest procedure see the following document:

http://dss.princeton.edu/training/Margins.pdf

The procedure using `prvalue` in the following pages does not work with Stata 13.
Predicted probabilities: using `prvalue`

After running `ologit` (or `logit`) you can use the command `prvalue` to estimate the probabilities for each event.

`Prvalue` is a user-written command, if you do not have it type `findit spost`, select `spost9_ado` from http://www.indiana.edu/~jslsoc/stata and click on “(click here to install)”

If you type `prvalue` without any option you will get the probabilities for each category when all independent values are set to their mean values.

```
. prvalue

ologit: Predictions for y_ordinal

Confidence intervals by delta method

          95% Conf. Interval
Pr(y=Agree|x): 0.3627 [ 0.3159, 0.4094]
Pr(y=Neutral|x): 0.2643 [ 0.2197, 0.3090]
Pr(y=Disagree|x): 0.3730 [ 0.3262, 0.4198]

   x1    x2    x3    x4    x5    x6    x7
   2.0020408  -8.914e-10  -1.620e-10  -1.212e-10   2.539e-09  -9.744e-10  -6.040e-10

Pr(y=Agree|x): 0.9916 [ 0.9738, 1.0094]
Pr(y=Neutral|x): 0.0055 [ -0.0061, 0.0172]
Pr(y=Disagree|x): 0.0029 [ -0.0033, 0.0090]
```

You can also estimate probabilities for a particular profile (type `help prvalue` for more details).

```
. prvalue, x(x1=1 x2=3 x3=0 x4=1 x5=2 x6=2 x7=4)

ologit: Predictions for y_ordinal

Confidence intervals by delta method

          95% Conf. Interval
Pr(y=Agree|x): 0.0029 [-0.0033, 0.0090]
Pr(y=Neutral|x): 0.0055 [-0.0061, 0.0172]
Pr(y=Disagree|x): 0.9916 [ 0.9738, 1.0094]

   x1   x2   x3   x4   x5   x6   x7
   1    3    0   -1    2    9    4
```

For more info go to: http://www.ats.ucla.edu/stat/stata/dae/probit.htm
Predicted probabilities: using `prvalue`

If you want to estimate the impact on the probability by changing values you can use the options `save` and `dif` (type `help prvalue` for more details)

```
Pr(y=Agree|x):   0.3730    0.3522    0.0208   [-0.0299,   0.0714]
Pr(y=Neutral|x): 0.2643    0.2641    0.0003   [-0.0026,   0.0031]
Pr(y=Disagree|x): 0.3627   0.3837   -0.0210   [-0.0737,   0.0317]
```

Here you can see the impact of x1 when it changes from 1 to 2.

For example, the probability of y=Agree goes from 35% to 37% when x1 changes from 1 to 2 (and all other independent variables are held at their constant mean values).

NOTE: You can do the same with logit or probit models
Useful links / Recommended books

- DSS Online Training Section [http://dss.princeton.edu/training/](http://dss.princeton.edu/training/)
- UCLA Resources to learn and use STATA [http://www.ats.ucla.edu/stat/stata/](http://www.ats.ucla.edu/stat/stata/)
- DSS help-sheets for STATA [http://dss/online_help/stats_packages/stata/stata.htm](http://dss/online_help/stats_packages/stata/stata.htm)
- *[Introduction to Stata*](http://fmwww.bc.edu/GStat/docs/StataIntro.pdf) (PDF), Christopher F. Baum, Boston College, USA. “A 67-page description of Stata, its key features and benefits, and other useful information.”
- Princeton DSS Libguides [http://libguides.princeton.edu/dss](http://libguides.princeton.edu/dss)

Books