

## Discussion Questions for Logistic Regression using Stata

Rather than go through my old notes in detail, I plan to have you discuss these questions in class (so spend some time working on them beforehand). Most of this is material you have learned in earlier courses, and/or can pick up from the web page notes on logistic regression. Not everyone has the same background, so anything that you don't understand you should be sure to ask about.

- Why do we use logistic regression, rather than OLS regression, in the first place? Talk about violations of OLS assumptions and plausibility of estimates.
- Consider the following output from three different logit commands:

```
. use "http://www.nd.edu/~rwilliam/xsoc73994/statafiles/logist.dta", clear
. logit grade gpa tuce psi
```

Iteration 0: log likelihood = -20.59173  
Iteration 1: log likelihood = -13.496795  
Iteration 2: log likelihood = -12.929188  
Iteration 3: log likelihood = -12.889941  
Iteration 4: log likelihood = -12.889633  
Iteration 5: log likelihood = -12.889633

```
Logistic regression                               Number of obs   =           32
                                                LR chi2(3)      =           15.40
                                                Prob > chi2     =           0.0015
Log likelihood = -12.889633                    Pseudo R2      =           0.3740
```

```
-----+-----
```

grade	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
gpa	2.826113	1.262941	2.24	0.025	.3507938 5.301432
tuce	.0951577	.1415542	0.67	0.501	-.1822835 .3725988
psi	2.378688	1.064564	2.23	0.025	.29218 4.465195
_cons	-13.02135	4.931325	-2.64	0.008	-22.68657 -3.35613

```
-----+-----
```

```
. logit, or level(99)
```

```
Logistic regression                               Number of obs   =           32
                                                LR chi2(3)      =           15.40
                                                Prob > chi2     =           0.0015
Log likelihood = -12.889633                    Pseudo R2      =           0.3740
```

```
-----+-----
```

grade	Odds Ratio	Std. Err.	z	P> z	[99% Conf. Interval]
gpa	16.87972	21.31809	2.24	0.025	.6524584 436.6945
tuce	1.099832	.1556859	0.67	0.501	.7637905 1.583721
psi	10.79073	11.48743	2.23	0.025	.6952747 167.4733
_cons	2.21e-06	.0000109	-2.64	0.008	6.74e-12 .7268044

```
-----+-----
```

```
. logit grade gpa tuce i.psi, nolog
```

```
Logistic regression                               Number of obs   =          32
                                                    LR chi2(3)      =          15.40
                                                    Prob > chi2     =          0.0015
Log likelihood = -12.889633                       Pseudo R2      =          0.3740
```

grade	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
gpa	2.826113	1.262941	2.24	0.025	.3507938 5.301432
tuce	.0951577	.1415542	0.67	0.501	-.1822835 .3725988
1.psi	2.378688	1.064564	2.23	0.025	.29218 4.465195
_cons	-13.02135	4.931325	-2.64	0.008	-22.68657 -3.35613

- In the above output, where is the model chi-square reported? What hypothesis does it test? If that number were not reported, how would you compute it based on the other information presented? [Hint: What do iteration 0 and iteration 5 tell you? And, as a sidelight, what is happening in all of these other iterations?]
- What is Pseudo R<sup>2</sup>? Is it the same as R<sup>2</sup> in OLS regression?
- How do you interpret the coefficients for gpa, tuce, etc? How does this differ from the interpretation of coefficients in OLS regression? What are some ways to make the results easier to interpret?
- What exactly did the 2<sup>nd</sup> logit command do? Why did it work? How was the 3<sup>rd</sup> logit command different from the first command? What does the i.varname notation accomplish?
- What are log odds, odds, odds ratios, and probabilities? How are they interrelated? How can they be easily computed in Stata?
- What is the difference between a Wald test and a Likelihood Ratio (LR) test? If I started with the command

```
logit y x1 x2 x3 x4
```

how would I use Wald and LR tests to test the hypothesis

H<sub>0</sub>:  $\beta_3 = \beta_4 = 0$

H<sub>A</sub>:  $\beta_3$  and/or  $\beta_4 \neq 0$

- What is the difference between these two Stata commands? When might each be appropriate?

```
test x3 x4
test x3 = x4
```

- What do we mean by a post-estimation command in Stata? What are examples of such commands?
- What is stepwise regression, and how do you do it in Stata? What are the advantages and disadvantages of stepwise regression?
- What is multicollinearity? How do you test for it in Stata when doing logistic regression?