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:

```
. 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
------------------------------------------------------------------------------
```

  - 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?]

  - What is Pseudo R^2? Is it the same as R^2 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? (NOTE: Much of the semester will be devoted to this last question, but you already have a few ideas presented in the notes.)

- What are log odds, odds 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_0: $\beta_3 = \beta_4 = 0$
  H_A: β_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 is multicollinearity? How do you test for it in Stata when doing logistic regression?

- What are "outliers?" What are some ways I can check for them when running logistic regressions?

Feel free to bring up any other questions that pertain to material in the notes.