ECON 407: Problem Set 4 3/26/2014

> This problem set is due Friday April 4 by 5pm. When responding to these questions take great care to interpret your results and tell me why you are doing what you are doing. A robot like stream of stata code and results will go some of the way to answering these questions but will not be sufficient without accompanying interpretation.
> As usual, please turn in a hard copy printout of your responses and incorporate Stata commands, results, and narrative in your written responses to this Problem Set. Combine these into a continuous narrative so I don't have to jump around looking at code and reading results.

Remember, to run `webuse set "http://rlhick.people.wm.edu/econ407/data"` before trying to webuse the data. To get the package of commands that includes `fitstat` and `prvalue`, issue the command `findit spost`. Click on the ado package corresponding to your version of stata, then select 'click here to install' on the right side of the page that pops up. If you've already installed `fitstat`, you can either use the command `ssc uninstall fitstat` or when stata tells you you've already installed it and gives you options under 'Possible things to do:' you can just select '4. Force installation replacing already-installed files' and the installation will continue.

1. Consider the random variates, x_i found in the spreadsheet here. You wish to model the mean of this distribution using the maximum likelihood approach for two statistical distributions: (1) the normal (this is exactly identical to what I did in class, and the sample spreadsheet will be quite helpful, so assume that $\sigma^2 = 1$) and (2) the poisson distribution. The probability density functions for these distributions are

   Normal PDF : $p(x_i|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x_i-\mu)^2}{2\times\sigma^2}}$

   Poisson PDF : $p(x_i|\lambda) = \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$

   Note: In the Poisson model, λ is the mean of the distribution.

   (a) Examine the pdf function values and see if they reside in the range [0,1], irrespective of the value of μ or λ you consider. Very briefly discuss how you did this.

   (b) Find the maximum likelihood estimate of the mean given each distribution. Report the maximum log-likelihood function value for each distribution.

   (c) Consider the Poisson Distribution. Using the likelihood ratio test, test $H_0 : \lambda = 1$ versus $H_0 : \lambda \neq 1$. Do you reject or fail to reject this NULL hypothesis?

   (d) Which is the preferred model for the mean: the normal or poisson MLE model?

2. Consider data from a contingent valuation study in Colorado accessed by `webuse southplatte`. The goal was to see if people were willing to pay for public works projects to restore recreational and ecosystem services on the South Platte River. Respondents were asked this question:

   > If the South Platte River Restoration Fund was on the ballot next election, and it cost your household \$__ each month in a higher water bill would you vote in favor or against?

   Where \$__ took on values of \$1, 2, 3, 8, 10, 12, 20, 30, 40, 50, 100 and was randomly assigned to each of the respondents (in this data, each row is a respondent). In the South Platte River Data, there are 95 observations on 8 variables. The variables are:

Table 1: South Platte River Data.

Variable	Description
YPAY	=1 if a yes response and =0 otherwise
BID	Increment to waterbill (randomly assigned)
UNLIMWAT	=1 if respondent thinks farmers are entitled to unlimited water, =0 otherwise
GOVTPUR	=1 if respondent believes the government should purchase land on the South Platte River, = 0 otherwise.
ENVIRON	=1 if a member of a conservation group
WATERBILL	Average waterbill for the community
HHINC	Before tax household income from all sources for 1997. There is a single missing value in the HHINC variable. For any model that uses HHINC, that observation is dropped.
URBAN	=1 if the respondent lives in a large city.

(a) Estimate a binomial logit, a probit, and an OLS model that models an individual saying 'yes' for this hypothetical referendum. Interpret your results and describe differences you find between the three models. Think carefully about how the variables *BID* and *WATERBILL* should enter the model. Compare marginal effects and parameter estimates.

(b) For the first 10 observations, calculate $p(\mathbf{x}_i\beta)$ and calculate whether the model predicts a yes or no vote. You may use either the probit or logit model for calculating the probability of a 'yes'. Contrast this with the actual vote given by each of the first 10 individuals. Provide a table and code generating your results. Using this information, provide a summary of predicted versus actual choices, analogous to stata's `lstat` command.

(c) For both models, discuss the predictive accuracy of the model by comparing the predicted outcomes with actual chosen outcomes, the (CountR2) statistic reported by the `fitstat` command. Contrast this with McFadden's R^2. Discuss what each type of "goodness of fit" measure is capturing.

(d) When comparing the probit and logit models, which do you recommend?

(e) Based on this data, suppose we wanted to predict the % of respondents who would vote 'yes' to this referendum if their current water bill was \$30 higher each month (note: respondents are currently *only* paying their current water bill and not current water bill + bid). Describe how the model parameters and probabilities can be used for this policy simulation. Being as specific as you can, investigate how the predicted probabilities for the first 10 observations differ from those calculated in part (b). What is your estimate of the % of 'yes' votes associated with this policy across all respondents?

(f) Describe how the matrix of second derivatives (also known as the information matrix) of the Likelihood Function, $I(\theta)^{-1} = \frac{\partial^2 lnL(\theta)}{\partial\theta\partial\theta'}$, is useful for calculating standard errors. Intuition and discussion, as opposed to formal proofs, is how I would approach this question.