Due: Feb. 15 by midnight

(1) Use the dataset "risky_behaviors.dta" which is saved in Canvas. The data is from a randomized trial targeting couples at high risk of HIV infection. The intervention provided counseling sessions regarding practices that could reduce their likelihood of contracting HIV. Couples were randomized either to a control group, a group in which just the woman participated, or a group in which both members of the couple participated. One of the outcomes examined after three months was "number of unprotected sex acts."

    (a) Model this outcome as a function of treatment assignment using a Poisson regression. Does the model fit well? Is there evidence of overdispersion?

    (b) Next extend the model to include pre-treatment measures of the outcome and the additional pre-treatment variables included in the dataset. Does the model fit well? Is there evidence of overdispersion?

    (c) Fit an overdispersed Poisson model. What do you conclude regarding effectiveness of the intervention?

    (d) These data include responses from both men and women from the participating couples. Does this give you any concern with regard to our modeling assumptions?

(2) Use the dataset "nes5200_processed_voters_realideo.dta" which is saved in Canvas to predict party identification (which is on a five-point scale) using ideology and demographics with an ordered multinomial logit model.

    (a) Summarize the parameter estimates numerically and also graphically.

    (b) Explain the results from the fitted model.

    (c) Use a binned residual plot to assess the fit of the model.

(3) Use the dataset "NSW.dw.obs.dta" which is saved in Canvas. Use the treatment indicator and pre-treatment variables to predict post-treatment (1978) earnings using a tobit model. Interpret the model coefficients.

(4) Use the dataset "oscars.csv" which is saved in Canvas. Also see "oscars.txt" for a description of the data. Fit a discrete choice model to predict winners of the Academy Awards.

    (a) Fit your own model to these data.

    (b) Display the fitted model on a plot that also shows the data.

    (c) Make a plot displaying the uncertainty in inferences from the fitted model.

(5) Use the dataset "hsbdemo.dta" which is saved in Canvas. The data set contains variables on 200 students. The outcome variable is "prog", program type. The predictor variables are social economic status, "ses", a three-level categorical variable and writing score, "write", a continuous variable. Entering high school students make program choices among general program, vocational program and academic program. Their choice might be modeled using their writing score and their social economic status.

(a) Fit a multinomial logistic regression model using social economic status and writing score to predict program type. Interpret the coefficients.

(b) Look at the averaged predicted probabilities for different values of the continuous predictor variable "write" within each level of "ses".

(c) Plot the predicted probabilities against the writing score by the level of "ses" for different levels of the outcome variable.