# Homework 9: GLM Introduction: Logit and Probit

Answer the following questions in a .pdf or .docx, explaining all of your answers and putting any tables and figures in the document as necessary. When data is called for to answer applied questions, I will provide it in bblearn. Turn in your R code that created all of the tables and figures separately, and be sure that it runs from source in such a way that it loads the data and performs all the tests without me fiddling with it. Make sure to document your R source code using # comments if you want partial credit.

For transformations, I will be asking to you perform and interpret generalized linear model regressions using traditional theory-based techniques for measures of uncertainty. Please include full tables with information I request in a proper regression table in your .docx or .pdf. For today, we will perform some regressions using some Gelman and Hill housing data in rodent.dta. I have also included a codebook for variable information for the file.

1. Theoretical Questions about Logit

   (a) Run a logit model with the independent variables `poverty_mean`, `regext`, `old`, `pubhous_mean`, `hispanic_mean`, `black_mean`, predicting the probability of a New York City apartment having rodent pests (`rodent2`). Then run a probit model with the same independent variables. Make a table with the results of the models side by side in the same table, making sure to include the coefficients (not odds ratios or risk statistics—use transform = NULL if using the tab_model command), standard errors, and 95% confidence interval.

   (b) Which variables do we conclude at the 99% confidence level affect the probability of a building having rodents and how do they affect the probability of rodents directionally?

   (c) Show the similarity in predictions over the values of poverty from the two models using R in two figures, graphically. Please be sure the Y axis to be the same on both graphs for comparability.

2. Interpretation of Coefficients

   (a) Estimate predicted probability from the logit model of having rodents for hypothetical apartments with specific characteristics who are otherwise average, then make a table displaying your results:

- An old building with 40% poverty in the neighborhood
- A 28% black neighborhood with regular exterminator service not in an old building.
- A rich neighborhood (`poverty_mean = .08`) with a neighborhood hispanic percentage of .35%

(b) Make a table showing the predicted probability of having rodents for percentages of public housing from 0 to 40% at every 5% (e.g. 0, 5%, 10%...) but being otherwise average.

(c) What is the marginal effect of a one percent change in poverty when poverty is at 11%, the building is not old, there is regular exterminator service, 9% of buildings are public housing, black household percentage is at 17%, and hispanic percent is 11%?