Gabriel Velez
Problem Set #1
MACS 30100

1) Classify a model from a journal

As I am a PhD student in Comparative Human Development, I am interested in how young people develop as citizens. With this in mind, I chose an article that discusses the relationship between a civic education program in the Dominican Republic and political participation. The statistical model I have selected comes from a generalized probit model that was then used to look at the effect of the civic education program (GAD below) on political participation.

a. Detailed citation:

Finkel, S. E. (2002). Civic education and the mobilization of political participation in developing democracies. *Journal of Politics*, *64*(4), 994-1020.

b. Mathematical or statistical model:

i. General model: Generalized Probit Model

$$y_i = \beta_k x_i + \beta_t T + E_i$$

Where yi represents political participation (a composite of four behaviors: taking part in community problem-solving activity; attending a local government meeting; working in an election campaign; and contacting a local elected official), x represents all independent variables, T represents treatment (this civic education program), B are the respective regression coefficients, and E_i is the error term.

ii. Full Model:

$$PolPart = \beta_1 Age + \beta_2 Age^2 + \beta_3 Gender + \beta_4 Edu + \beta_5 Inc + \beta_6 Emp + \beta_7 Church + \beta_8 City + \beta_9 GroupMem + \beta_{10} PolInt + \beta_{11} MediaUse + \beta_{12} Vot96 + \beta_{13} GAD + E_i$$

Where PolPart is the measure of the individual's political participation (0-4 scale, and the dependent variable of interest), Age is age category (with 5 groups; Age^2 is squared), Gender is binary with 1=male, Edu is the highest education level completed by the individual, Inc is the yearly income for the individual's household (7 categories), Emp is whether the individual is employed (1/0 dummy), Church is a measure of the individual's religiosity (on a five point scale), City is a dummy for whether they live in a city, GroupMem is the proportion of voluntary organizations the individual belongs to out of possible 10, PolInt is a measure of individual's interest (on a 1-4 scale), MediaUse is a measure of the individual's exposure to popular media (on a 1-4 scale), Vot96 is whether the individual voted in the 1996 elections (1=yes), and GAD is whether the individual participated in the civic education program.

c. **Exogenous Variables (taken as given from outside the model):** Age, Age^2, Gender, Education, Income (though one could argue that it may be determined by Education and Employed), Employed, Church, City, GroupMem, PolInt, MediaUse, Vot06, and GAD

d. **Endogenous Variables:** PolPart

e. Classifying model:

i. **Static/dynamic:** This model is static because it does not vary over time (that is, does not account for time-dependent changes, but measures political participation in equilibrium). Though Vot96 is from a previous time, it is not a repeated or changing measure over time.

ii. **Linear/nonlinear:** The model is linear because all of the individual parameters are linear. Even though it has a squared term (Age^2), Age is still a linear parameter.

iii. Deterministic/stochastic: The model is stochastic because it has an error term to define the randomness that exists across individuals (i.e., it is a random variable).

f. Variable or feature possibly missing: There is no variable to account for the cultural context of political participation that encompasses the individual. In other words, this model does not take into account the local community's political participation. This could be accounted for, by example, by including a variable as well on the proportion of voting eligible residents of the community (possibly defined by a political district) who voted in the 1996 election.

2. Make your own model

a. Model of whether or not someone decides to get married (at least one endogenous variables must be 1=get married, 0=not get married)

$$Mar = \beta_1 Age + \beta_2 Inc + \beta_3 PartInc + \beta_4 Urban + \beta_5 Religion + \beta_6 ParMar + \beta_7 MarStatus + \beta_8 DatingStatus + \beta_9 DatingLength + \beta_{10} Age * DatingLength + \beta_{11} PrevMar$$

Where Mar is the endogenous variable of whether or not someone gets married, Age is the age of the person, Inc is the individual's income level, PartInc is their partner's income level, Urban is where or not they live in an urban area (defined by population density, and as a 1 or 0 variable), Religion is a categorical variable of different religious affiliations for the individual in question, ParMar is the length of the individual's parents' marriage (0 if never married), MarStatus is whether or not the person is currently married, DatingStatus is what the current status of the individual, DatingLength is the length of the current relationship (if there is one), Age*DatingLength is an interaction term, and PrevMar is a binary of whether or not the person has been married before..

a. What are the key factors and why did you decide on those factors and not others?

I hypothesize that the key factors would be Religion, Dating Status, and Dating Length. First, I would say Religion because cultural context and expectations could be significantly related to whether someone gets married (in terms of social aspects like whether friends are married and family/community norms). While it may be hard to directly measure that, Religion could serve as a proxy since certain religions may place greater importance on marriage and those who are involved in a religious group may be more likely to have friends from that same group (thus adding to the social pressure). Dating Status and Dating Length would be important because although people may marry when single and without a partner (or being with a partner for a short time), there is probably a higher likelihood that if they have had a partner for a certain amount of time that they will get married. I also included an interaction term of Age and Dating Length that I believe may be a key factor because at younger ages, dating for a longer time may be less correlated with getting married (e.g., a 20 year-old who has been dating someone for 2 years may be less likely to get married than a 38 year-old who has been). I feel less certain about this being important because the correlation could be positive or negative (maybe those who are older are more likely to be fine being in a long-term committed relationship without marriage), but it seems like length of dating would certainly differentially relate to probability of getting married based on age.

The other factors may be important, and are also related to characteristics about the individual (like income and ParMar), their partner (partner's income), and their context (Urban), but I think these are all factors that have more noise and may not be as highly correlated. For example, people across all income brackets get married, but there is some evidence that the rates are higher at certain income brackets.

b. How could you do a preliminary test whether factors are significant in real life?

To explore whether these factors are significant, I would first try to use available demographic databases to test the model. Due to some of the factors, however, I might have to use survey methods. The Census data, for example, might provide some of the information I am looking for (like age, income bracket, urban, religion) and I could use this by taking people who had been married in the previous six months or year (if that is available on the census) and running the model. I do not believe, however, that the census (or similar datasets) would have all of the explanatory variables (such as Dating Status or ParMar) and may not even include how recently someone was married (which could be used to test the model because most of the factors would have been consistent for the time period right before they got married as well as the current time period). In this case, I would most likely have to recruit a random sample to survey. In the survey, I could ask about each one of my explanatory variables and then do a follow up six months later to see if the individual did in fact get married. Once I have collected enough data to have sufficient power for the analyses, I could test whether the predicted key factors are in fact important. Specifically, I could run a regression model with and without key factors and then see how they differ (using R-squared values and F statistics to compare the explanatory power of the models and exploring in greater depth any factors I did not list as key ones but that were significant).