**Education 265: Problem Set 6**

**Try to do the problem set entirely on your own (but feel free to ask for help from other students or us).**

**Due:** Friday, February 12, 2020 at 1:00 p.m. in the class Canvas folder (Assignments tab 🡪 Problem Set 6). Turn in

* Your answer sheet as a Word document (titled “**LastName\_PS6**”)
* Type your answers in **bold directly** under the questions
* Your do-file (titled “**LastName\_PS6\_Code.do**”).

**Objectives:**

1. We will practice estimating dichotomous outcome variables using logistic regression and estimating marginal probabilities.

**Set up your do-file.** Start by setting the path directory (use the *cd* command) so that any data files are saved to the same folder. Call the data into Stata (*use for\_ps6.dta, clear*).

1. On last problem set we predicted the probability of being rated as high on externalizing behaviors (“highext”) using zmath1 and our set of controls below:

* standardized fall of kindergarten reading achievement (‘zread1’);
* dummies for Blacks, Hispanics, Asians, and Other ethnicity (omit Whites as the comparison group);
* dummy for mothers with a high school diploma, vocational school, or some college; and dummy for mothers that graduated from college or attended graduate school (omit mothers with less than a high school diploma as the comparison group);
* and the Fryer and Levitt variables, namely: age in month, birthweight in ounces, female, number of children’s books, number of children’s books squared, mother over 30 at first birth, mother a teenager at first birth (mother age 19-29 at first birth omitted as reference group), continuous measure of socioeconomic status, mother received WIC benefits.

Restricted to our analysis sample again to sample\_ps5==1

Create a global using the following code (same code from last PS)

global controls zread1 dblack dhispanic dasian dother ///

ageinmonth bw\_oz dfemale books books2 dmage\_30plus dmage\_teen wksesl momwic ///

dhighcol dcolplus

* 1. (This is the same as problem set 5’s question #8) Run a linear probability model (LPM) in which you predict ‘highext’ using fall of kindergarten math score, gender, and our set of controls. Interpret the coefficient produced by zmath1. (don’t forget to restrict your sample!)

**Coefficient: -.037 (.006) at p-value < .001**

**A 1 standard deviation increase in fall of kindergarten math score is associated with a 3.7 percentage point decrease in probability of getting a higher that mean score in spring kindergarten externalizing, while holding control variables constant.**

* 1. Now run the same model, but use the “logit” command instead of “regress.” Interpret the coefficient produced by zmath1 in terms of 1) log odds and 2) change in odds.

**Change in log odds: -.165**

**Odds ratio: e^(-.165) = .848**

**At p-value < .001**

**A 1 standard deviation increase in fall of kindergarten math score is associated with a .165 decrease in log odds or a 15% decrease in odds of getting a higher that mean score in spring kindergarten externalizing, while holding control variables constant.**

* 1. Run the post-estimation “margins” command to estimate the *probability* of being considered high externalizing for standardized math scores 1 standard deviation below, 1 standard deviation above, and at the mean (all other variables should be held at their means). Interpret these three numbers.

**.4595 (.009) at p-value < .001**

**The probability of being considered high externalizing for students 1 standard deviation below mean in fall of kindergarten math score is 45.95%, holding all other predictors at their grand mean.**

**.4189 (.004) at p-value < .001**

**The probability of being considered high externalizing for students at mean in fall of kindergarten math score is 41.89%, holding all other predictors at their grand mean.**

**.3794 (.007) at p-value < .001**

**The probability of being considered high externalizing for students 1 standard deviation above mean in fall of kindergarten math score is 37.94%, holding all other predictors at their grand mean.**

**Note: Mean of zmath1 variable is 2.15e-10 so it’s small enough to be considered approximately zero.**

* 1. Run 1b again, then run the post-estimation “margins” command *and option dydx* to estimate the *change in probability* of being considered high externalizing for a change at standardized math scores 1 standard deviation below, 1 standard deviation above, and at the mean (all other variables should be held at their means). Interpret these three numbers.

**-.041 (.007) at p-value < .001**

**The decrease in probability of being considered high externalizing** **for 1 standard deviation increase in fall of kindergarten math score is 4.1 percentage points for students who are 1 standard deviation below the mean, holding all other predictors at their grand mean.**

**-.040 (.007) at p-value < .001**

**The decrease in probability of being considered high externalizing** **for 1 standard deviation increase in fall of kindergarten math score is 4 percentage points for students who are at the mean, holding all other predictors at their grand mean.**

**-.039 (.006) at p-value < .001**

**The decrease in probability of being considered high externalizing for 1 standard deviation increase in fall of kindergarten math score is 3.9 percentage points for students who are 1 standard deviation above the mean, holding all other predictors at their grand mean.**

Time you took to finish this problem set: \_\_\_\_2\_\_