### Education 265: Problem Set 3

**Due:** Friday, January 22, 2020 at 1:00 p.m. in the class Canvas folder (Assignments tab à Problem Set 3). Turn in

* Your answer sheet as a Word document (titled “**LastName\_PS3**”)
* Type your answers in **bold directly** under the questions
* Your do-file (titled “**LastName\_PS3\_Code.do**”).

**Objectives:** In this problem set, we will continue to clean our class dataset and create variables that will be useful for running analyses. Specifically, we will investigate the relation between ethnicity and kindergarten teacher-reported internalizing, externalizing, and approaches to learning. We will generate a descriptive statistics table, run regression models, and interpret the coefficients (FUN, FUN, FUN!). The tables and figures created for this exercise should be formatted according to APA standards (e.g., <https://owl.english.purdue.edu/owl/resource/560/01/>).

1. Set up your do-file. Start by setting the path directory (use globals or just the cd command) so that any data files are saved to the same folder. Call the data into Stata (*use* ***for\_ps3.dta****, clear*).
2. Locate the variables for teacher-reported internalizing, externalizing, and approaches to learning. We will use the fall AND spring kindergarten measures of these behavioral ratings (*Hint: use “lookfor” command to find the variables. Beware, our dataset also contains parent-reported ratings of the same behaviors. The teacher-reported variables should begin with the letter “t” (e.g., t1learn, t1extern, etc.).* 
   1. On each of the 6 variables, missing values need to be recoded to “.” (*Hint: after recoding them, you should have a mean of “2.96” for teacher-reported approaches to learning in the fall, with a range of 1 to 4).* Recode these values and paste your code into your answer sheet.

**forvalues i = 1(1)2 {**

**replace t`i'intern = . if t`i'intern < 0**

**replace t`i'extern = . if t`i'extern < 0**

**replace t`i'learn = . if t`i'learn < 0**

**}**

1. **Descriptive Table**
   1. Set an analysis sample “*ps3\_sample*” that only includes students who had non-missing data on all 6 teacher behavioral reports, and non-missing data on race/ethnicity. The analysis sample should include 16,873 students. Paste your code for creating this analysis sample into your answer sheet.

**gen ps3\_sample = 0**

**replace ps3\_sample = 1 if !mi(t1intern, t2intern, t1extern, t2extern, t1learn, t2learn, race)**

**tab ps3\_sample**

* 1. Generate a single descriptive table that includes means, standard deviations, and values at the 25th and 75th percentiles for each of our 6 behavioral measures. Present these descriptive statistics for the full analysis sample, White students in the analysis sample, Black students in the analysis sample, and Hispanic students in the analysis sample. All descriptive statistics should include two digits after the decimal. Format the table according to APA standards and guidelines outlined in the lab; see below for an example. Paste the table into your answer sheet. (My table is in blue, and your table should be in black)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1 Descriptive statistics of behavioral measures | | | | | | | | | | | | | | | | |
|  | Full sample |  |  |  | White |  |  |  | Black |  |  |  | Hispanic |  |  |  |
|  | mean | sd | p25 | p75 | mean | sd | p25 | p75 | mean | sd | p25 | p75 | mean | sd | p25 | p75 |
| Internalizing Problem - fall K | 1.54 | 0.53 | 1.25 | 2 | 1.52 | 0.52 | 1.25 | 1.75 | 1.57 | 0.56 | 1 | 2 | 1.52 | 0.52 | 1.25 | 1.75 |
| Internalizing Problem - spring K | 1.58 | 0.52 | 1.25 | 2 | 1.56 | 0.52 | 1.25 | 2 | 1.63 | 0.57 | 1.25 | 2 | 1.56 | 0.52 | 1.25 | 2 |
| externalizing Problem - fall K | 1.62 | 0.63 | 1 | 2 | 1.6 | 0.62 | 1 | 2 | 1.79 | 0.7 | 1.2 | 2.2 | 1.6 | 0.62 | 1 | 2 |
| externalizing Problem - spring K | 1.68 | 0.65 | 1.2 | 2 | 1.64 | 0.63 | 1 | 2 | 1.9 | 0.75 | 1.2 | 2.2 | 1.64 | 0.63 | 1 | 2 |
| Approach to Learning - fall K | 2.98 | 0.67 | 2.5 | 3.5 | 3.05 | 0.66 | 2.5 | 3.67 | 2.8 | 0.71 | 2.17 | 3.33 | 3.05 | 0.66 | 2.5 | 3.67 |
| Approach to Learning - spring K | 3.1 | 0.69 | 2.67 | 3.67 | 3.17 | 0.67 | 2.67 | 3.83 | 2.9 | 0.74 | 2.33 | 3.5 | 3.17 | 0.67 | 2.67 | 3.83 |
| Observations | 16873 |  |  |  | 9514 |  |  |  | 2449 |  |  |  | 9514 |  |  |  |
| SD = Standard Deviation, p25 = 25th percentile, p75 = 75th percentile | | | | | | | | | | | | | | | | |

1. **Regression Analysis** 
   1. Run ANOVA comparing ethnicity group differences (i.e. Black, Asian, Hispanic, White, Other; *hint: use simplified race categorical variable*) on spring teacher-reported externalizing, internalizing, and approaches to learning. Paste your code below.

**oneway t2extern race\_simp if ps3\_sample == 1**

**oneway t2intern race\_simp if ps3\_sample == 1**

**oneway t2learn race\_simp if ps3\_sample == 1**

* 1. What is the mean externalizing behavioral problems for White students and Asian Students? What is their group difference in externalizing behavioral problems? Is this group difference statistically significant? How do you know (report the apprpriate statistics)? **1.63, 1.47**

**-0.16 (0.02)**

**Since the P-value for their pairwise anova is very low, the results are statistically significant.**

* 1. Run three regression models investigating the links between ethnicity (i.e., race dummies) and spring teacher-reported externalizing, internalizing, and approaches to learning. The dummy variable for White will be omitted as the comparison group. Paste the code for all three regressions into your answer sheet.

**foreach var of varlist t2extern t2intern t2learn {**

**reg `var' dblack dhispanic dasian dother if ps3\_sample == 1**

**}**

* 1. Report and interpret the coefficient reporting the relationship between whether a student identifies as Asian and teacher reported externalizing behavioral problems in spring of kindergarten. How do these results compare to the ANOVA results? **-0.16 (0.02)**

**Asian students have 0.16 lower teacher reported externalizing behavioral problems compared to white students on average in the spring of kindergarten.**

**The results from ANOVA (asian vs white) are the same as asian stats in the regression results.**

* 1. What is the average score on the spring approaches to learning measure for white students in the analysis sample? How did you figure it out from the regression output from step 4.c? **3.16**

**Since we are using white students as a reference here in the regression, the constant value in the regression (b0) would be the average score on the spring approaches to learning measure for white students.**

* 1. Standardize the fall and spring teacher reported measures of internalizing, externalizing, and approaches to learning. Copy and paste your code for this step into your answer sheet.

**foreach var of varlist t1intern t2intern t1extern t2extern t1learn t2learn {**

**egen z`var' = std(`var')**

**}**

* 1. Re-run the regression models from step 4.c with the standardized dependent variables. Did the results meaningfully change when compared with the raw score models?

**Units have changed in standardazied variables, therefore coefficients now show that how much a 1 standard deviation change in independent variables can change in units of SD in the dependent variable. So the results have changed in that sense.**

* 1. Report and interpret the coefficient relating the standardized measure of spring teacher-reported internalizing to whether a student identifies as Black. **0.134 (0.022)**

**Black students have 0.134 units of standard deviation more internalizing problems compared to white students on average.**

* 1. Now, run the set of standardized regressions again, but add the lagged measure of the dependent variable for each model as a control. Paste your code into the answer sheet (*Hint: This means that for the externalzing regression, the dependent variable will be the spring externalizing measure and the independent variables will be the set of ethnicity dummies, in addition to the standardized measure of fall teacher-reported externalizing).*

***reg zt2extern dblack dhispanic dasian dother zt1extern if ps3\_sample == 1***

***reg zt2intern dblack dhispanic dasian dother zt1intern if ps3\_sample == 1***

***reg zt2learn dblack dhispanic dasian dother zt1learn if ps3\_sample == 1***

* 1. Compare the internalizing model in step 4.h with the model in step 4.i (i.e., the standardized model without lagged control vs. the standardized model with lagged control). Describe what happened to the coefficients produced by the set of ethnicity dummies once the control for fall internalizing was added. Intuitively, what does this change suggest about the nature of teacher’s perceptions of student behavioral problems over the course of the kindergarten year?

**Adding the control varibales helps us to understand the effect of race on the dependent variable better, because it cancels out the impact fall results had on the race. By adding controls on fall variable, we can start to compare students of other ethnicities with white students considering both have the same fall results with just helps us to single out the effect of race/ehtnincity.**

1. **Regression Table**
   1. Generate a table that presents estimates of the relations between ethnicity and spring kinderharten teacher-reported beahviors when 1) not controlling for fall teacher ratings ratings and 2) controlling for fall teacher ratings. The table should include 2 models for each spring measure of behavior (i.e., internalizing, externalizing, and approches to learning): the first model will just include the set of ethnicity dummies and the second model will add the lagged depenent variable. Omit White students as the comparison group in each of the models.

In other words, generate an APA-formatted table that displays the estimates from steps 4.h and 4.i. Your table should have 6 columns in total, and include notes on anything needed to fully interpret the coefficients. These notes are very important because you want your table to be self-contained as much as possible so that a reader who is skimming your paper can fully understand your table. Paste the table into your answer sheet (*Hint: remember to use your analysis sample variable to keep the sample size constant across the models)*.

|  |  |  |  |
| --- | --- | --- | --- |
| Table 2. Behavioral measurements gap between white students and other ethnicities in spring of kindergarten | | | |
|  |  |  |  |  |  |  |
|  | (1) | (2) | (3) | (4) | (5) | (6) |
|  | externalizing behavior | internalizing behavior | approach-to-learning | externalizing behavior | internalizing behavior | approach-to-learning |
|  |  |  |  |  |  |  |
| black | 0.40\*\*\* | 0.13\*\*\* | -0.38\*\*\* | 0.18\*\*\* | 0.08\*\*\* | -0.13\*\*\* |
|  | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) |
|  |  |  |  |  |  |  |
| hispanic | 0.03 | 0.07\*\* | -0.19\*\*\* | 0.02 | 0.02 | -0.03\* |
|  | (0.02) | (0.02) | (0.02) | (0.01) | (0.02) | (0.02) |
|  |  |  |  |  |  |  |
| asian | -0.25\*\*\* | -0.16\*\*\* | 0.22\*\*\* | -0.10\*\*\* | -0.12\*\*\* | 0.15\*\*\* |
|  | (0.03) | (0.03) | (0.03) | (0.02) | (0.03) | (0.02) |
|  |  |  |  |  |  |  |
| other | 0.20\*\*\* | 0.17\*\*\* | -0.23\*\*\* | 0.07\*\* | 0.10\*\*\* | -0.06\* |
|  | (0.03) | (0.03) | (0.03) | (0.02) | (0.03) | (0.02) |
|  |  |  |  |  |  |  |
| Standardized externalizing Problem - fall K |  |  |  | 0.72\*\*\* |  |  |
|  |  |  |  | (0.01) |  |  |
|  |  |  |  |  |  |  |
| Standardized Internalizing Problem - fall K |  |  |  |  | 0.56\*\*\* |  |
|  |  |  |  |  | (0.01) |  |
|  |  |  |  |  |  |  |
| Standardized Approach to Learning - fall K |  |  |  |  |  | 0.70\*\*\* |
|  |  |  |  |  |  | (0.01) |
|  |  |  |  |  |  |  |
| Constant | -0.06\*\*\* | -0.03\*\* | 0.09\*\*\* | -0.01 | -0.00 | 0.00 |
|  | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
|  |  |  |  |  |  |  |
| Observations | 16873 | 16873 | 16873 | 16873 | 16873 | 16873 |
| R-squared | 0.026 | 0.006 | 0.026 | 0.527 | 0.311 | 0.501 |
|  |  |  |  |  |  |  |
| Standard errors in parentheses |  |  |  |  |  |  |
| Note: First three columns include regression results without controlling for fall measurements and the second three columns include results with controlling of fall measurements. Whites measurements are omitted from the set  therefore, all the coefficients are relative to that group" | | | | |
| \* p<0.05 | \*\* p<0.01 | \*\*\* p<0.001" |  |  |  |  |

1. Save your dataset as “for\_lab4.dta”. Amount of time it took you to finish this problem set: \_\_8hrs\_\_\_\_\_
2. Now that we are in week 3, do you have any feedback for lecture or for lab? Is there anything particualry confusing? Anything you want us to cover? Any command you think would be ideal to learn at this point?

**A feedback I can give on problem sets is that the questions are actually very detailed and a lot of my time is spent making sure if I am using the right variable or is this the corrent number or not.**

1. **Bonus:** (this is a hard one!) Generate a bar graph in Stata that displays means of standardized externalizing behaviors, internalizing behaviors, and approaches to learning by race (i.e., estimates from 4.e). White students should be included on your graph. The graph should include 3 sets of bars: one for each dependent variable. Paste the graph and code into your answer sheet. (*Hint: One easier way to create this graph is to copy and paste the coefficients into a clean data set. You can also use the “postfile” command. Use “graph bar” with the “over” option) Just know that you cannot simply just graph from the stored dataset that you saved at 7.*

This is an example of what it would look like (yours might be a little different): 