Data Mining: Final Project

David Fraire

5/16/2019

Effects of Head Start on PPVT at Age 3, College Enrollment, & Highschool Graduation

The following regressions show in table form will represent the effects of the Headstart program on the outcome variables of **PPVTat3**, **somecollege**, **hsgrad**. We decided to use a mixture of linear models and logit models to proceed with analysis of the effect of Headstart program on the three aforementioned dependent variables. The table showing effects of Head Start on **PPVTat3** is a linear model. The tables showing Head Start effects on **somecollege** and **hsgrad** are logit models and the shown coefficients have been exponentiated for interpretation reasons.

Table 1: Headstart effect on PPVT Scores at Age 3

	Dependent variable:			
	PPVTat3			
	(1)	(2)	(3)	(4)
headstart	−6.741*** (1.054)	−6.739*** (1.055)	−2.992*** (1.010)	−3.440** (1.677)
Hispanic			−8.504*** (1.073)	−8.274*** (1.173)
Black			−12.205*** (0.946)	−13.120*** (1.066)
Male		0.030 (0.852)	0.193 (0.783)	0.702 (0.777)
hsgrad				3.046*** (0.785)
FirstBorn				3.721*** (0.784)
headstart:Black				3.125 (2.245)
headstart:Hispanic				−1.380 (2.717)
Constant	25.028*** (0.477)	25.013*** (0.645)	29.140*** (0.671)	25.691*** (0.875)
Observations	984	984	984	984
R^2	0.040	0.040	0.191	0.228
Adjusted R^2	0.039	0.038	0.188	0.222
Residual Std. Error	13.348 (df = 982)	13.355 (df = 981)	12.270 (df = 979)	12.014 (df = 975)
Note:				*p<0.1; **p<0.05; ***p<0.01

For the previous table, we begin in the first by regressing the data for **PPVTat3** only on **headstart** using a linear fit. The effect of the headstart program in this model shows a very statistically significant negative effect of -6.741. The linear model in the last column includes **Hispanic**, **Black**, **Male**, **hsgrad**, **FirstBorn**, and interaction terms for **headstart**. The very significant variables from this regression are **headstart**, **Hispanic**, **Black**, **Male**, **hsgrad**, and **FirstBorn**. The **Black** variable has a very large average effect on the outcome of PPVT scores and is the highest in magnitude. The **headstart** variable still has a significant effect but is lower in magnitude and lower in signiicance than that of previous regressions with less covariates. This means that participation in Headstart has a consistently negative effect on the test scores with the data provided in our dataset even after controlling for other highly significant variables. Given that the children who participate in Headstart are likely to come from disadvatanged socioeconomic backgrounds, this can be explained in the following models as those who are in Headstart will likely have lower scores to start. The last model included variables that were shown to explain a significant amount of variation in the original dataset and were therfore chosen to be included in order to convey the decreasing marginal effect of Headstart on PPVT scores at age 3.

Table 2: Headstart effect on College Enrollment

	Dependent variable:			
	somecollege			
	(1)	(2)	(3)	(4)
headstart	1.625*** (0.055)	1.637*** (0.056)	1.308*** (0.058)	0.841*** (0.130)
Male		0.740*** (0.045)	0.737*** (0.045)	0.679*** (0.050)
Black			1.997*** (0.053)	1.982*** (0.069)
Hispanic			1.619*** (0.059)	1.649*** (0.073)
LogInc_0to3				1.145*** (0.032)
headstart:Black				1.491*** (0.158)
headstart:Hispanic				1.249*** (0.186)
Constant	0.275*** (0.025)	0.318*** (0.032)	0.243*** (0.040)	0.112 (0.344)
Observations	11,470	11,470	11,470	7,485
Log Likelihood	−6,162.935	−6,140.100	−6,049.910	−4,664.154
Akaike Inf. Crit.	12,329.870	12,286.200	12,109.820	9,344.308
Note:				*p<0.1; **p<0.05; ***p<0.01

For the effect of Head Start on college enrollment which is measured by the binary variable **somecollege**, we used a logit model to see the effects on the odds of having gone to college. Note, that the coefficients in the table have been exponentiated for ease of interpretation. Simply running a regression of **headstart** on **somecollege** shows that **headstart** has a significant effect on the outcome of having enrolled at a college. The odds of attending college are increased by a factor of 1.625. However, this effect is shown to decrease and even decrease the odds of attending college when controlling for other significant variables such as **Male**, **Black**, **Hispanic**, **LogInc_0to3**, and interaction **headstart** variables. In the final column with all included covariates, the odds of going to some college are .841 if you participated in the **headstart** program and is very significant. This tells us that after controlling for other highly significant variables, that attending the Head Start program actually hurt your odds of attending some college. Once again, this can be explained by the participants of Head Start already having lower chances of pertaining to an environment where college acheivement is lower than that of a more affluent child.

Table 3: Headstart effects on High School Graduation

	Dependent variable:			
	hsgrad			
	(1)	(2)	(3)	(4)
headstart	1.625*** (0.055)	1.637*** (0.056)	1.308*** (0.058)	0.854*** (0.130)
Male		0.740*** (0.045)	0.737*** (0.045)	0.677*** (0.050)
Black			1.997*** (0.053)	1.933*** (0.069)
Hispanic			1.619*** (0.059)	1.695*** (0.074)
LogInc_0to3				1.088*** (0.035)
MothED				1.044*** (0.011)
headstart:Black				1.473*** (0.158)
headstart:Hispanic				1.229*** (0.186)
Constant	0.275*** (0.025)	0.318*** (0.032)	0.243*** (0.040)	0.112 (0.342)
Observations	11,470	11,470	11,470	7,479
Log Likelihood	−6,162.935	−6,140.100	−6,049.910	−4,650.817
Akaike Inf. Crit.	12,329.870	12,286.200	12,109.820	9,319.634
Note:				*p<0.1; **p<0.05; ***p<0.01

In order to measure the effect of Head Start on a child graduating high school we decided to use a logit model in order to tell us the odds ratio. To start, we regressed **headstart** on **somecollege** and then exponentiated the coefficient for the all the regressions included in the table in order to tell us the effect on the odds of receiving the outcome of having gone to college for particular children. When regressing **headstart** only on **somecollege** we see that being enrolled in the program increases the odds of going to college by a factor 1.625 and is very significant. Our last column includes the variables **Male**, **Black**,**Hispanic**,**LogInc_0to3**,**MothED**, and interaction **headstart** terms. Once again, the effect of **headstart** improves the odds of graduating highschool by a factor of 1.625 and is statistically significant. Once we add for other important and highly significant variables, the **headstart** variable decreases the odds of graduating high school. We believe that this is caused by selection bias within our dataset as the effects in Deming's original research paper also showed this interesting result on high school graduation.