Hidden Curriculum Assignment

Patrick Massey

02/18/2022

Data

I use a subset of data from the 1997 National Longitudinal Survey of Youth, which contains longitudinal data on individuals from 1997 to 2019. In particular I use observations from 2002 that focus on the incarceration status of an individual for that year. When analyzing the data I remove any individuals for which there is no incarceration information available, or if they began the year already incarcerated. After cleaning the data I am left with 8,621 observations. I then create an dummy indicator value for if the individual is incarcerated at all during the year. This indicator variable *incarcerated* will be the variable of interest for the model.

Empirical Analysis

In this analysis I seek to estimate the probability of an individual becoming incarcerated based on race and sex. Looking at Figure 1 below, we see that Black Males have the highest incarceration rate of almost 6%. Males in general have a higher incarceration rate than Females. One anomaly is for the Mixed Race (Non-Hispanic) group. This should be seen as an small sample size issue because out of the 8,621 observations Mixed Race (Non-Hispanic) make up 81 observations.

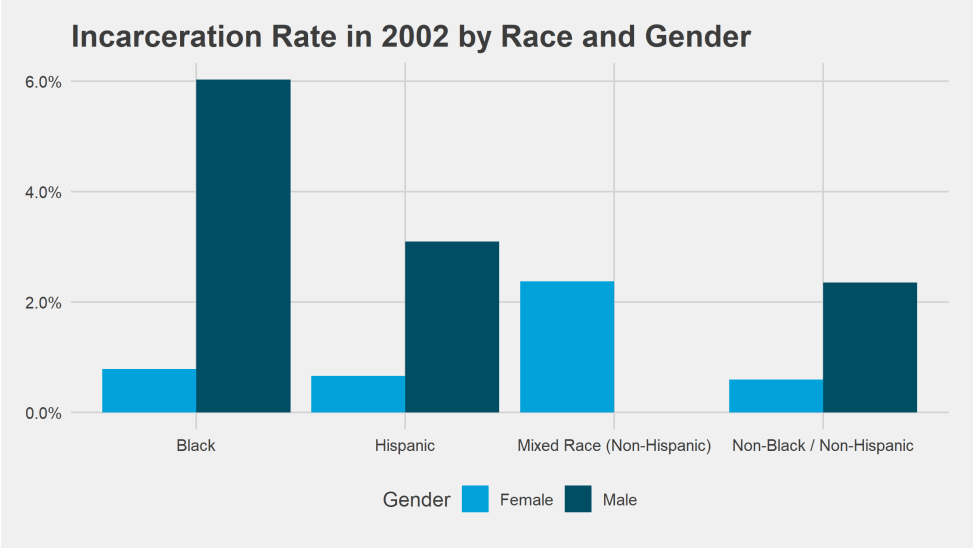

Figure 1: Incarceration Rate in 2002 by Race and Gender

I summarize the graph in Table 1 below which shows the incarceration rate broken down by sex and race.

Table 1: Incarceration Rate in 2002 by Race and Gender

Gender	Black	Hispanic	Mixed Race Non Hispanic	Non Black Non Hispanic
Female	0.0079225	0.0066225	0.0238095	0.0059798
Male	0.0602740	0.0309498	0.0000000	0.0235602

The model we are seeking to estimate is shown below:

$$Y_i = \beta X_i$$

Where Y_i is a binary value for an individuals incarceration status and X_i is a vector containing sex and race characteristics. I summarize the regression results in Table 2 below.

Table 2: Regression Output. Omitted category is Black Females.

	Dependent variable:
	Incarcerations in 2002
Hispanic	-0.015^{***}
	(0.005)
Mixed Race (Non-Hispanic)	-0.021
	(0.013)
Non-Black / Non-Hispanic	-0.019^{***}
	(0.004)
Male	0.028^{***}
	(0.003)
Constant	0.020^{***}
	(0.003)
Observations	8,621
R^2	0.012
Adjusted R^2	0.012
Residual Std. Error	0.141 (df = 8616)
F Statistic	27.193^{***} (df = 4; 8616)
Note:	*p<0.1; **p<0.05; ***p<0.01

All variables are statistically significant at the 99% level except for Mixed Race (Non-Hispanic) category which I think goes back to the small n issue.