My Paper on NLSY97 Data

Jacob Bulzak

February 2022

1   Introduction

In this report we examine patterns in incarceration status by race and gender using the NLSY97 data from the year 2002. The analysis involved examining the incarceration status of individuals in the dataset and tracking the mean number of months incarcerated for the year 2002 based on race and gender. We note that the average number of months incarcerated tends to be very low, as the majority of survey respondents were not incarcerated at all for the year 2002.

We run the following regression:

$$y = \beta_1 + \beta_2 Hispanic + \beta_3 MixedRace(Non-Hispanic) + \beta_4 Non-Black/Non-Hispanic + \beta_5 Male + \varepsilon$$
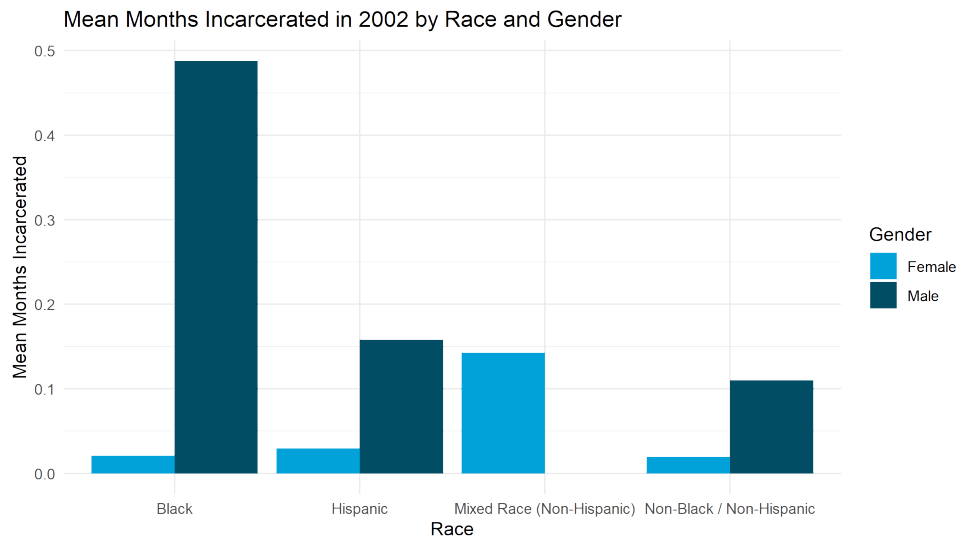
2   Analysis

Figure 1: Mean Number of Months Incarcerated in 2002 by Race and Gender

The above graph shows the mean number of incarcerations in 2002 by race and broken down by gender. In every category, (except for Mixed Race (Non-Hispanic)) the mean number of months incarcerated is higher for males than for females. In fact, Mixed Race Non-Hispanic males have a mean incarceration time of zero months. Finally, we observe that the average duration of incarceration for Black males is much higher than for any other category.

Table 1: Mean Months Incarcerated in 2002 by Race and Gender

Gender	Black	Hispanic	Mixed Race Non Hispanic	Non Black Non Hispanic
Female	0.0211268	0.0298013	0.1428571	0.0193192
Male	0.4876712	0.1579509	0.0000000	0.1099476

Table 1 serves as numeric representation of Figure 1. The table reiterates the results seen in the graph above: Black males exhibit the maximum mean incarceration time, whereas Mixed Race Non-Hispanic males zero months incarcerated in 2002, on average.

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

	Dependent variable:
	Months Incarcerated in 2002
Hispanic	−0.159***
	(0.038)
Mixed Race (Non-Hispanic)	−0.174**
	(0.083)
Non-Black / Non-Hispanic	−0.189***
	(0.035)
Male	0.194***
	(0.022)
Constant	0.155***
	(0.026)
Observations	8,621
R^2	0.015
Adjusted R^2	0.014
Residual Std. Error	1.019 (df = 8616)
F Statistic	32.033*** (df = 4; 8616)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 2 summarizes the results of a the regression introduced in Section 1 that relates race and gender to incarceration time. We note that the model's coefficient of determination (i.e. R-squared) is 0.015. This indicates only a very small proportion of the variation in incarceration time is predictable from our covariates, implying a poor model fit. Conversely, each of the p-values associated with the covariate coefficients is significant at the 1 percent level. This tell us that changes in race and gender are indeed associated with changes in incarceration time, and furthermore, that these covariates belong in our regression. The table also demonstrates being Black is the greatest contributor to incarceration time given that it is the omitted variable in our regression and all other races have negative coefficients associated with them. Specifically, being a black male corresponds to, on average, a 0.194-month increase in time incarcerated. Finally, we note that the regression intercept (constant) represents the average number of months incarcerated for a Black female.