Report on NLSY97 Incarceration Data

Amal Kadri

Sping 2022

1 Introduction

Below is an analysis of the 1997-2019 NLS Youth Data on `Incarceration`. Most of the data set is fairly good, but unfortunately there were very few observations for `Mixed Race` individuals in the data set, so most of the results I obtain aren't very useful for that subcategory of individuals, as can be seen below.

2 Section 1: Graphs

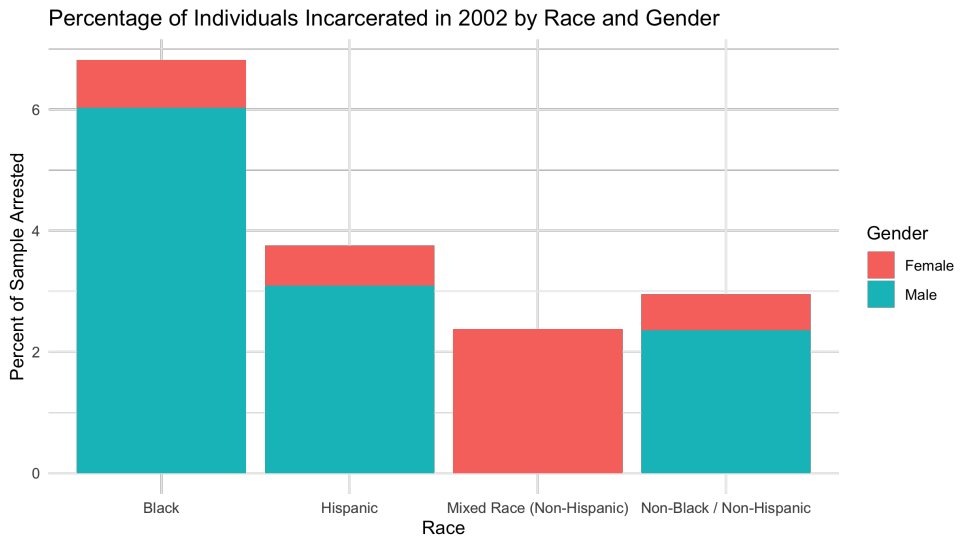

Figure 1: Percentage of Individuals in the Survey who were incarcerated at any point in 2002, broken down by race and gender. Mixed Race proportions should be largely ignored due to small sample size.

3 Section 2: Table

Table 1: Percentage Incarcerated in 2002 by Race and Gender

Gender	Black	Hispanic	Mixed Race Non Hispanic	Non Black Non Hispanic
Female	0.7922535	0.6622517	2.380952	0.5979761
Male	6.0273973	3.0949840	0.000000	2.3560209

This table shows a breakdown of the percentage of the individuals in the survey that were incarcerated, broken down by `Race` and `Gender`. The percentages for the `Mixed Race (Non-Hispanic)` category because the sample size for that subcategory is so small (81 individuals, only 1 female of which was incarcerated). That aside, there is a clear trend toward men being incarcerated, with `Black Men` being almost 8 times as likely to be incarcerated as `Black Women`. Black individuals seem to be more likely to be incarcerated than other races regardless of gender.

4 Section 3: Regression

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

	Dependent variable:
	Likelihood of Being Incarcerated in 2002
Hispanic	-0.618^{***}
	(0.209)
Mixed Race (Non-Hispanic)	-1.021
	(1.036)
Non-Black / Non-Hispanic	-0.866^{***}
	(0.172)
Male	1.660^{***}
	(0.205)
Constant	-4.470^{***}
	(0.197)
Observations	8,621
Log Likelihood	-810.209
Akaike Inf. Crit.	1,630.419
Note:	*p<0.1; **p<0.05; ***p<0.01

The trend from above continues in the regression table. Computed using a Logit Regression, with Incarceration Status being regressed on `Race` and `Gender`. Because the coefficients of the Log-Likelihood function are somewhat difficult to interperet (the model takes the form: $p/(1-p) = e_0^{\beta} + e_1^{\beta} x_1$), the magnitude of beta coefficients in the regression output may seem off. For example, the odds of a `Black Male` being incarcerated are $e^1.66 = 5.26$ times more likely than a `Black Female`. Similarly, all other races have lower odds of being incarcerated than Blacks (again mostly ignoring the Mixed Race (Non-Hispanic) category for sample-size reasons).