Analysis of 2002 Incarceration data from NLSY97

Peter (Tun-Shuo) Lee

Sping 2022

1 Introduction

Systematic inequity between racial groups in the US has always been a widely debated topic. Differences in employment, education, income, have placed many families in a disadvantage that spans across generations. Amongst the racial treatment discussions, perhaps one of most contentious is of the criminal justice system and whether their representatives judge the cases differently based on the defendants' ethnicity. This paper will analyze the racial composition of the `National Longitudinal Study of Youth`'s data from 2002, in particular, the racial and gender distribution of the incarcerated. While this short analysis aims not to provide a definitive conclusion to the ongoing debate, it does provide quantitative evidence on the difference (or lack there of) of incarcerations in youth across race and gender.

2 Analysis

The data set NLSY97 included 8621 individuals' gender and race information, as well as whether they were incarcerated during each month in 2002. By first summing the number of months, and dividing by the total number of individuals within each gender and race group, we essentially find the rate of incarceration of the youths for each race and gender. The data is summarized in the table below:

Table 1: Mean incarcerations 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

An interesting trend arises in that males have a generally higher incarceration rate than females. This trend is present for all races, except for Mixed Race NonHispanic, where there are no recorded male incarcerations. Our findings can be better interpretted once visualized:
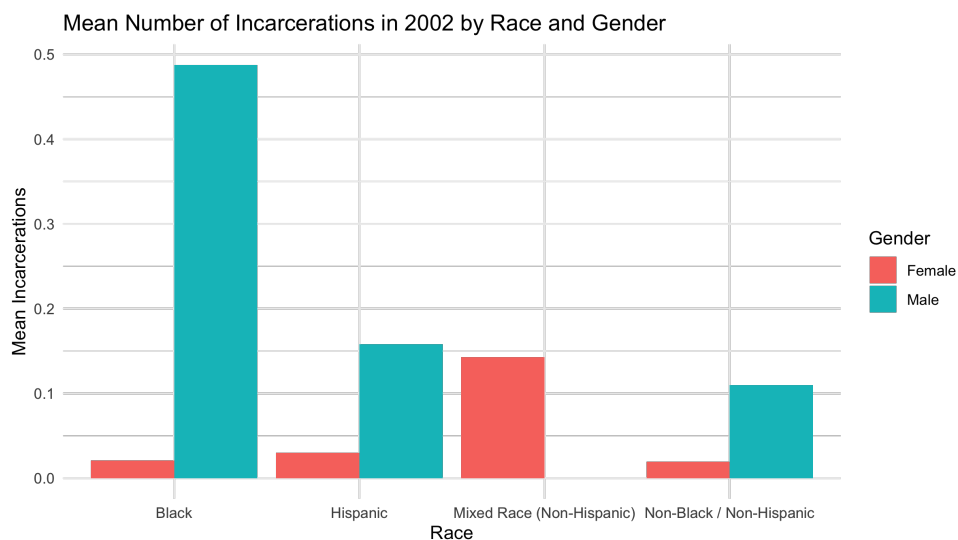

Figure 1: Mean Number of Incarcerations in 2002 by Race and Gender (this is the LaTeX caption, not the ggplot title)

For Black, Hispanic and Non-Black Non-Hispanic individuals, the incarceration rate for male is higher than female. The difference is the largest for Black males, then Hispanic, and the lowest being NonHispanic, NonBlack. With both male and female considered, Black individuals are the most likely to be incarcerated. Hispanic teens comes next, and lastly, NonBlack/ NonHispanic.

Surprisingly, the incarceration rate for Mixed Raced, NonHispanic women are higher than for women of all other races, and yet, there are no recorded incarcerations for Mixed Race, NonHispanic men.

3   Regression

I regressed the incarceration rate on categorical variables representing race and gender. I omitted Black Females to avoid multicolinearity.

$$incarcerationrate = \beta_0 + \beta_1 Hispanic + \beta_2 MixedRace(NonHispanic) + \beta_3 NonBlack/NonHispanic + \beta_4 Male + \varepsilon$$

The results are as follows:

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

	Dependent variable:
	Arrests 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