

My Paper on NLSY97 Data

Yu Xia

Spring 2022

1 Introduction

This paper talks about incarceration status of different gender and race. In order to examine the difference between different categories, this paper uses the data provided by the National Longitudinal Survey. I chose NLSY97, which interviewed many American youth in 1997. There is a lot of information about education, employment, marriage, etc. This paper focuses on the incarceration status in the year 2002.

Firstly, I filter out wrong data. Incarceration data of previous months are dropped as well. Thus, all incarceration I'm looking at happened in 2002. Thereafter, I categorise the data by race and gender, according to the original data. Meanwhile, I take the sum of each person's incarceration variable. It shows how many months this person spent incarcerated. For example, If the sum is 9, this person spent 9 months incarcerated. Then I take the mean time of incarceration by each group. As shown below, this paper compares different categories by figure, simple econometric model and tables.

The econometric model used in this paper is:

$$Incarceration = Hispanic\beta_1 + MixedRace\beta_2 + NonBlackAndNonHispanic\beta_3 + Male\beta_4 + \varepsilon$$

where *Incarceration* means average incarceration time in the year 2002, *MixedRace* is the binary variable of whether or not the person is non-hispanic mixed race, *NonBlackAndNonHispanic* is the binary of whether or not this person is neither black nor hispanic, *Male* stands for the binary variable for sex.

2 Interpretation

The figure below shows the mean incarceration time of different gender and sex.

In this data, none of mixed race (non-hispanic) male were incarcerated in 2002. So the mean incarceration time for them is 0. For all other races, the mean incarceration time for male is higher than for females.

From the perspective of comparing races, black male have a very long incarceration time on average, but black females' time is relatively low. Mixed race non-hispanic females have a pretty long incarceration time on average. The rank of male's mean incarceration time is different from female, showing that there are some other factors affecting the incarceration time. But it is beyond the scope of the paper.

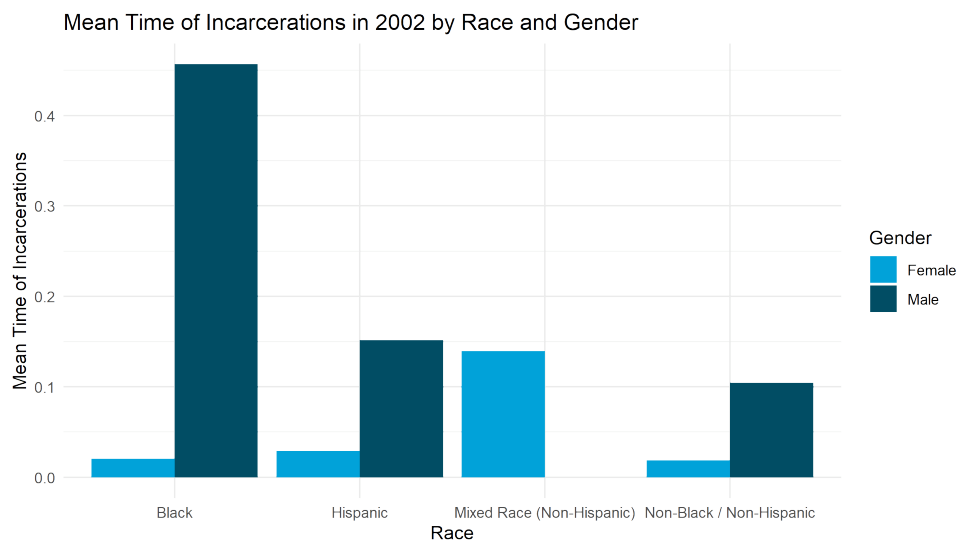


Figure 1: Mean Time of Incarceration in 2002 by Race and Gender

Table 1: Mean incarceration time in 2002 by Race and Gender

Gender	Black	Hispanic	Mixed Race Non Hispanic	Non Black Non Hispanic
Female	0.0205832	0.0292208	0.1395349	0.0186501
Male	0.4568007	0.1514841	0.0000000	0.1044343

Table 1 shows the mean incarceration time categorised by race and gender. Comparing different categories shows the same result as Figure 1. However, these numbers can be given meaning in reality. For example, the mean incarceration time for black females, in this data set, is about 0.02 month. Other categories have similar intuition. These numbers are pretty small, it is because

only a little part of interviewee were incarcerated in 2002. But these numbers can still be used for comparison, just like above.

Table 2 shows the regression result of a simple metrics model.

The coefficient of "Mixed Race (Non-Hispanic)" is significant at 0.1 level, and all other coefficients are significant at 0.01 level.

These coefficients can provide us with some intuition. Take the coefficient of Hispanic as an example. -0.149 means that holding all else equal, compared with Black Females, the Hispanic Females spent 0.149 month less in incarceration in expectation. For Hispanic Males, holding all else equal, they spent $-0.149+0.183=0.034$ more months incarcerated in expectation. Here's what the coefficient of "Male" stands for. Other categories can be interpreted accordingly.

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

	<i>Dependent variable:</i>
	Incarcerations in 2002
Hispanic	-0.149*** (0.037)
Mixed Race (Non-Hispanic)	-0.163** (0.081)
Non-Black / Non-Hispanic	-0.179*** (0.033)
Male	0.183*** (0.021)
Constant	0.148*** (0.024)
Observations	8,984
R ²	0.014
Adjusted R ²	0.013
Residual Std. Error	0.999 (df = 8979)
F Statistic	31.172*** (df = 4; 8979)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01