

My Paper on NLSY02 Data

Aff Mazhar

Spring 2022

1 Data

We will be analyzing the dataset of National Longitudinal Survey of Youth from 1997. The data set reflects incarceration rates from the year 2002 and the respective gender/race associated with incarceration.

2 Analysis

In order to provide qualitative analysis over the dataset given, included is three visualizations that depict the data uniquely. Figure 1 portrays the mean number of incarcerations per respective demographic in the year 2002. Figure 2 quantifies the previous visualization, and Figure 3 is a regressed form of the entire dataset.



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

As displayed above, the largest mean of incarcerations is Black Male with an average of 0.49 (reflected numerically below in Table 1). Non-black/Non-Hispanic females are the higher mean comparative to the men in the respective category with a 0.107 average. Mixed Race (Non-Hispanic) males are non-existent in Figure 1. Table 2 depicts a regression output of the dataset where Hispanic, Mixed Race, and the Non-Black categories are have a negative coefficient. This indicates that there is a negative relationship with the increase in the number of incarcerations and the respective category depicted over 8,978 observations. The Male category, however, shows a positive relationship with incarcerations. This indicates as the amount of incarcerations increase by 0.01, there is a 0.194 increase in the amount of males incarcerated. The r-squared value is very low so this regression does not provide perfect, accurate measures of the regression at hand.

Table 1: Mean incarcerations in 2002 by Race and Gender

Gender	Black	Hispanic	Mixed Race Non Hispanic	Non Black Non Hispanic
Female	0.0206009	0.0292524	0.1395349	0.0240000
Male	0.4905822	0.1709314	0.0000000	0.1073798

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

	<i>Dependent variable:</i>
	Incarcerations in 2002
Hispanic	−0.156*** (0.039)
Mixed Race (Non-Hispanic)	−0.180** (0.081)
Non-Black / Non-Hispanic	−0.192*** (0.034)
Male	0.194*** (0.022)
Constant	0.159*** (0.025)
Observations	8,978
R ²	0.014
Adjusted R ²	0.014
Residual Std. Error	1.050 (df = 8973)
F Statistic	32.087*** (df = 4; 8973)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01