

# Causal Inference Hidden Curriculum Assignment

Scott Kjorlien

Spring 2022

## 1 Summary Statistics

For this NLSY97 dataset of incarcerated youths, I chose to analyze the average number of months incarcerated by Race and Gender, given some level of incarceration (that is, sum of all monthly indicators is greater than 0).

Figure 1 shows a simple bar plot of the mean number of months incarcerated in 2002. We can see that that black men have the highest number of months incarcerated on average, while black women have the lowest. The gender difference between black men and black women appears striking, but By reducing the dataset to only incarcerated individuals, and subgrouping by both race and gender, I'm greatly reducing the number of individuals of each subgroup. Therefore, large variation between groups could easily be a matter of small sample size, rather than a representation of a group's sentencing.



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

Table 1 shows the numbers associated with each bar in Figure 1. As we see in Table 2, once we select only individuals who were incarcerated, we have limited data on the Mixed Race Non Hispanic group, and generally less data points for females.

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

Gender	Black	Hispanic	Mixed Race Non Hispanic	Non Black Non Hispanic
Female	2.666667	4.500000	6	3.230769
Male	8.090909	5.103448	NA	4.666667

Table 2: Sample Size of Each Group

Gender	Black	Hispanic	Mixed Race Non Hispanic	Non Black Non Hispanic
Female	9	6	1	13
Male	66	29	NA	54

## 2 Regression Summary

The results from our regression below show that incarcerated black males serve on average 2.61 months more time than incarcerated black females. Further, incarcerated non-black/Non-Hispanic females serve on average 2.8 months fewer than black females. In order to analyze this further across different groups, I would want to test different linear combinations of these groups. Surprisingly, even with the limited sample size, we can have quite a few statistically significant coefficients which we can use for inference.

Table 3: Regression Output. Omitted category is Black Females.

	<i>Dependent variable:</i>
	Months Incarcerated in 2002
Hispanic	-2.306*** (0.824)
Mixed Race (Non-Hispanic)	0.857 (0.729)
Non-Black / Non-Hispanic	-2.859*** (0.658)
Male	2.610*** (0.741)
Constant	5.143*** (0.729)
Observations	178
R <sup>2</sup>	0.161
Adjusted R <sup>2</sup>	0.142
Residual Std. Error	3.946 (df = 173)
F Statistic	8.302*** (df = 4; 173)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01