

Exploratory Data Analysis

Anastasiia Saenko. Antara Bhide. Daniela Jiménez Lara

Data Overview:

This report pretends to answer either of the following research questions:

- **How does the type of authority making the arrest and their use of violence influence an inmate's perception for the reason for their incarceration in Mexico?**
- **How does the existence of extortion from authorities to inmates during the arrest influence an inmate's sentence in Mexico?**

For this purpose, we are using the Mexican National Survey of Incarcerated Population (EN-POL) implemented by the Mexican National Institute of Statistics and Geography (INEGI) during 2021. Its general objective is to provide information on the experience of the criminal procedure and internment of the incarcerated population in Mexico aged 18 and over. The survey was designed to produce statistically relevant results and estimations at the national level, state level and prison level.

The general themes covered by the survey are the following:

- Sociodemographic and socioeconomic characteristics
- Family and employment background
- Legal and criminal background
- Judicial procedure
- Arrest information
- Experience with the district attorney's office
- Experience in the penitentiary center
- Exit expectations.

To carry out the survey, 203 penitentiary centers were visited between June and July of 2021, 54 of them were of interest. The survey size contemplated ***67584 people aged 18 and over, of which 55,535 are men and 12,049 women. However, only the complete surveys were published in the final dataset, hence the dataset contains 61,499**** rows, each representing an inmate that was interviewed. Each column represents the answer to a question or follow-up answers for the same questions. The dataset is divided into seven csv files containing the 11 modules of the survey with a total of 1286 columns.

- Source: INEGI (2021). *Encuesta Nacional de Personas Privadas de a Libertad*. [Data set]. INEGI. <https://www.inegi.org.mx/programas/enpol/2021/>.

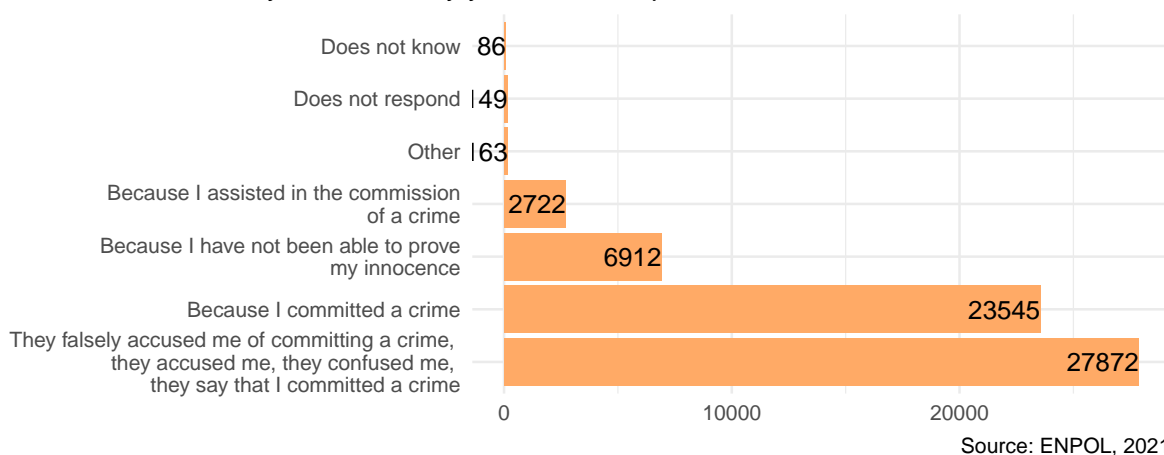
Module	Content
Module 1	sociodemographic characteristics: nationality, marital status, education, sexual orientation
Module 2	Housing and home condition prior to arrest, work history and economic income
Module 3	Information related to the context or circumstances in which the informant's arrest occurred
Module 4	Experiences of inmates during their stay at the Public Ministry Agency (D.A.office), as well as t
Module 5	Information related to the crimes (allegedly) committed, characteristics of the criminal process,
Module 6	Information about the characteristics of the prison cells, goods and services provided by the pris
Module 7	Lifestyle within the Penitentiary Center: sports, recreation, religious worship). Family visits rece
Module 8	Possible acts of corruption associated with the provision and access to different goods and servic
Module 9	Criminal and family history of inmates
Module 10	Exit expectations of inmates and self-perception
Module 11	Condition of the interview, that is, situations that arose during the development of the interview

Outcome variables:

Research question 1: How does the type of authority making the arrest and their use of violence influence an inmate's perception for the reason for their incarceration in Mexico?

The outcome variable for this research question is an inmate's perception for the reason of their incarceration, measured in question 3.1 of the ENPOL survey: Could you tell me why you are in this prison?

Question 3.1. Could you tell me why you are in this prison?

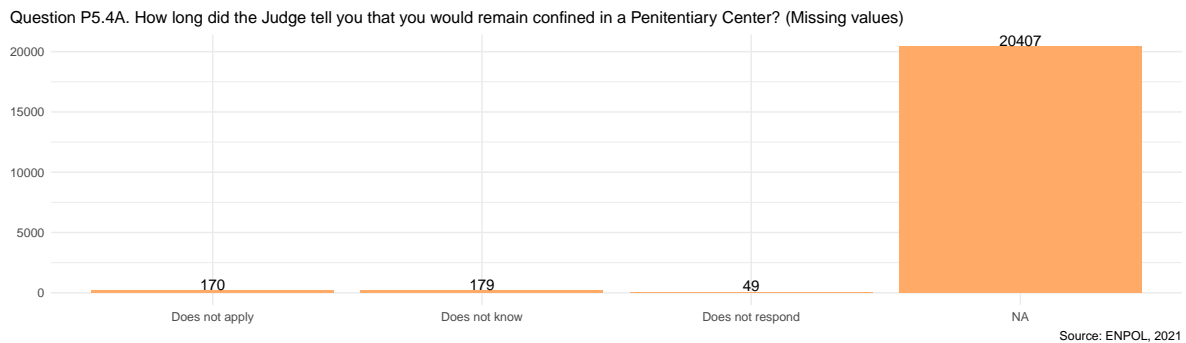
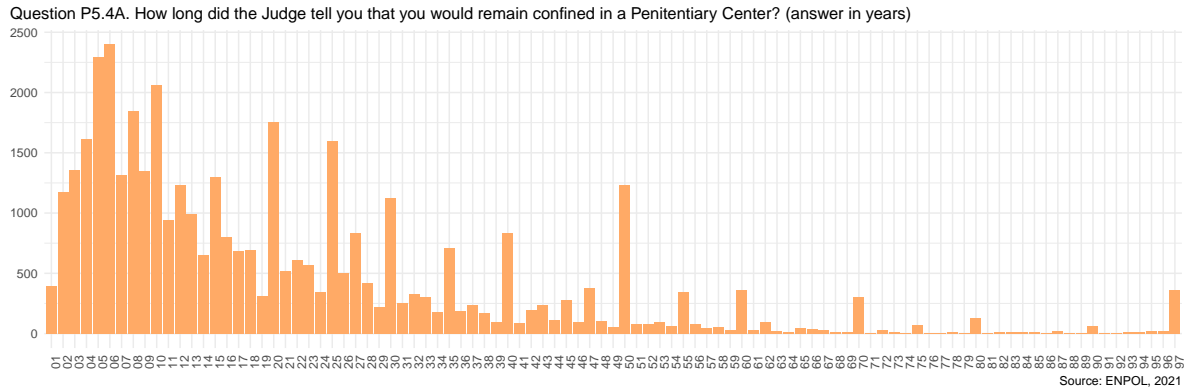


The distribution of the outcome variable tells us that 42% (26,267) of the inmates perceive that the reason for their incarceration is because they committed a crime or assisted in the commission of one. However 56% (34,784) perceive that they are in prison because they have

not been able to prove their innocence or because they have been falsely accused. For this outcome variable, the missing values are pretty low with .3% of no answers.

Research question 2: How does the existence of extortion from authorities to inmates during the arrest influence an inmate's sentence in Mexico?

The outcome variable for this research question is the inmate's years in their prison sentence, measured in question 5.4 of the ENPOL Survey: How long did the Judge tell you that you would remain confined in a Penitentiary Center?



For this outcome variable, the missing values plus the no-answer rate account for around 33% of the data. The distribution of the years in prison sentences does not follow a normal distribution, though most of the data concentrates in 1 to 20 years of prison sentence.

Primary relationships of interest:

Present descriptive statistics and exploratory plots in whichever format you think is best (tables, figures) for your primary relationship of interest (dependent variable and primary independent variable, if applicable). Describe your findings. Whether you have one primary independent variable of interest or multiple will depend on your research question.

##Other characteristics:

Briefly describe other variables in the data. If there are many, do not list them all. Rather, describe the types of variables that are present (e.g., “demographic information”).

##Potential challenges:

Describe aspects of the data that may present challenges in the modeling stage. For example, might certain categorical variables need to be collapsed? Do you have any missingness, particularly in key variables of interest? Could the size of the dataset present model selection challenges?