

Background

In January 2012, the Cook County State's Attorney's Office established a program intended to reduce re-arrest among people on bail awaiting trial. The program ran through October 2013. The SA's Office asked you to evaluate the effectiveness of the program and provided the data described below. The person who created the program left, so we know little about how people were assigned to the program, but we do know assignment was done independently at each arrest. This means that it's possible that someone arrested multiple times could have been referred to the program on some occasions but not on others.

Data

There are three data sets: case.csv, demo.csv, and prior_arrests.csv. Case.csv is the main dataset showing dates of arrest and disposition (trial or court appearance) during the time the intervention program ran, as well as an indicator of whether the arrestee was referred to the intervention program for that arrest. Demo.csv contains some demographic information about the arrestee, and prior_arrests.csv contains information on individuals' arrest histories from 2008-2011.

Dataset Descriptions

	N	Level
Case.csv	25000	Case (one row per case)
Demo.csv	14353	Person (one row per person)
Prior_arrests.csv	43602	Arrest (one row per arrest)

Case.csv

	Caseid	Person_id	Arrest_date	Dispos_date	Treat
Type	Numeric	Numeric	Character	Character	Numeric
Range	(4, 100000)	(1, 20000)	2012-01-02, 2013-10-28	2012-01-05, 2018-03-04	(0, 1)
Description	A unique identifying number for the arrest	A unique identifying number for the individual	Arrest Date	Date of Trial or Court Appearance	1 = treated, 0 = not treated

Demo.csv

	Person_id	race	gender	bdate
Type	Numeric	Character	Character	Character
Range	(1, 20000)	ASIAN, BLACK, HISPANIC, WHITE	M = male, F = Female	Birth Date
Description	A unique identifying number for the individual	The individual's racial category	The individual's gender	The individual's date of birth

Prior_arrests.csv

	Person_id	Arrest_date
Type	Numeric	Character
Range	(1, 20000)	2008-01-01, 2011-12-31
Description	A unique identifying number for the individual	Arrest Date

Part 1: Data Management

- The main outcome of interest is whether the defendant was arrested again prior to his/her disposition (trial or court appearance) date. Create a variable called 're_arrest' that equals one if the defendant was arrested prior to trial and zero otherwise. E.g., imagine person 1 was arrested three times.

Caseid	Person_id	Arrest_date	Dispos_date	Re_arrest
123	1	2012-01-01	2012-04-05	0
1344	1	2012-06-05	2012-09-13	1
1451	1	2012-07-09	2012-11-21	0

- Create a variable called prior_arrests that equals the number of arrests prior to the current arrest. Assume that all of the individual's arrests prior to the study period are contained in prior.csv. For example, if person z has five arrests in prior.csv and two in case.csv, her first arrest in case.csv should have prior_arrests = 5 and the second should have prior_arrests = 6. If someone is not included in prior_arrests.csv, assume they had zero arrests at the start of the study period.
- Create an 'age' variable that equals the defendant's age at the time of each arrest.

Part 2: Statistical Analysis

Help the SA's Office determine if the program should be continued/expanded by estimating the program's effect on re-arrests prior to disposition. To draw conclusions about this program's effect, answer the following questions.

- Describe the study population based on the data available to you.
- Are the treatment and control groups balanced, or are differences (race, gender, age, and prior arrests) in the composition of the two groups?
- Did participating in the program reduce the likelihood of re-arrest before disposition? Explain your answer and your methodology.