Instructions: Please write/type your answers NEATLY and hand in a hard copy before class starts. If you choose to write the solution by hand, please write in legible writing and do not try to fit answers in weird spaces.

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.

Data

There are two data sets: case.csv and demo.csv. Case.csv is the main dataset showing dates of arrest and disposition (trial or court appearance) during the time the intervention program ran. The file also contains an indicator of whether the arrestee was referred to the intervention program for that arrest, whether the person was re-arrested while awaiting trial, and the number of prior arrests at the time of the current arrest. Demo.csv contains some demographic information about the arrestee.

Dataset Descriptions

File Name	N	Level
Case.csv	25000	Case (one row per case)
Demo.csv	14353	Person (one row per person)

Case.csv

	Caseid	**Person_id**	**Arrest_date**	**Dispos_date**	**Treat**	**Re_arrest**	**Prior_arrest**
Type	Numeric	Numeric	Character	Character	Numeric	Numeric	Numeric
Range	(4, 100000)	(1, 20000)	2012-01-02, 2013-10-28	2012-01-05, 2018-03-04	(0, 1)	(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	1= re-arrested while awaiting trial, 0 =	Number of arrests prior to this cases

						not re-arrested	arrest_date

Demo.csv

	Person_id	**race**	**gender**	**bdate**
Type	Numeric	Character	Character	Character
Range	(1, 20000)	ASIAN; BLACK, HISPANIC; BLACK, NONHISPANIC; WHITE, HISPANIC, WHITE NONHISPANIC	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

Part 1: Data Management

1. Convert the excel files into Stata data sets.
2. The demographic data was extracted from a system that inconsistently coded gender. Recode it so that males are consistently coded as "M" and females are consistently coded as "F".

3. The demographic data also combines race and ethnicity into a single variable. Recode this variable so that Race contains three values (ASIAN, BLACK, WHITE), and a new variable called ethnicity has two values (HISPANIC, NONHISPANIC).

4. Merge the Case and demo datasets together so that each row in the case dataset also contains the demographics of the defendant.

5. Create an 'age' variable that equals the defendant's age at the time of the case's arrest date.

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.

1) Describe the demographic characteristics of the study population based on the data available to you. (hint: the study population has 25,000 subjects).

2) Are the treatment and control groups balanced (on race, gender, etc.), or are their differences in the composition of the two groups?

3) Did participating in the program reduce the likelihood of re-arrest before disposition? Explain your answer and your methodology.

Deliverables Expected:

- A do-file which answers Part 1 and also contains code for later parts (added as an appendix to the write up).
- A Write up for part 2 (document)
- A Log File for the analysis (Optional)
- The regression output in a document file (Optional)