Article-Questions

Mackenzie Fry

October 23, 2017

# Question 1:

The unit of observation in the Brady and Li study in this study is the person we know this because that is the file of FARS data that we are choosing from. That is also the dataset we want to use to replicate the study results we need to choose the DBF files from years 1999 to 2010, and the person file because that is our unit of observation.

# Question 2:

The constraints that are used for the data we need, include the years 1999-2010, the six states selected (California, Hawaii, Illinois, New Hampshire, Rhode Island, West Virginia. They also mention they considered all ages and box sexes. They focused on individuals who had died (fatal injury) within 1 hour after the crash (time constraint). They looked at drug tests of alcohol and other drugs whether the tests were positive or negative. We also want to get rid of values for age, drug and alcohol, and other variables with missing values because they specifically mention in their methods section excluding these values.

We want to filter the variables down to st\_case, veh\_no, per\_no, per\_type, lag\_hrs, lag\_mins, inj\_sev, age, alc\_res, sex, drug\_res of any kind which is simply just drug results. We want to keep st\_case, veh\_no, as is but more specifically unique identifier values for the observation of person variables. For lag\_hrs we want to only keep 0 and 1 hrs, anything more we don’t want. For lag\_mins we want 60 or less because that indicates one hour. For inj\_severe we want to keep injuries with a value of 4, because that indicates a fatal injury. Per\_type is equal to 1 which is the driver because we want to analyze if drivers are under the influence or not. All ages were included, for states, we want to get specifically 6, 15, 17, 33, 44 and 54 which are the six states we want. The alc\_res variable value should be 94 which indicates .94 or greater. We want to use drug\_res to understand what kind of drug they used

# Question 3:

The variables we would want to use to get the stratified variables (grouping by) would be age, year, sex, alc\_res, drug\_res originally and then later year category by 1999-2002, 2003-2005, and 2007 to 2010. But we would want to manipulate the data before grouping by these variables. In the end, we would want to group by age category which is later broken down, sex, blood alcohol level calculated later, and the drug type which is later mutated by different values to different drug types.

Age Category is coded as the number of their age. 0 is less than one year old, and 999 indicates a missing value for age. (we likely won’t see any ages below 16 because we are focused on the driver and you must be 16 to drive).

Year is broken down by the year of the data set (1999- 2010) Later we will category year by 1992-2002, 2003-2005, and 2007-2010 as the authors did.

Alcohol\_res is later calculated to be blood alcohol level, but originally they use to the actual blood level 0-93, 94 indicates .94 or greater, in 1999-2009 a 95 indicated test refused but after 2009 it was just left blank or --, in 2010 a 95 indicated the blood alcohol level is not reported, 96, or 996 indicates the test was not given. A 97 or 997 indicates that the AC test was performed but the results are unknown, a 99, 999 means the test was either unknown or not reported. There was a change in indication over the years so 96 and 97 later changed to 997 or 996, and 999 for missing. We are calculating in the end, whether there was alcohol or no alcohol (depsite what the blood alocohol level is, for example we are not interested if they were over the limit .08)

drug\_res is broken down by type 100-295 indicates narcotic, 300-395 depressant, 400-495 is a stimulant, 500-595 is a hallucinogen, 600-695 cannabinoid, and 800- 895 is anabolic steroids, 900-995 is an inhalant. 999 indicates unknown or not reported a score of 1 classified no drugs or negative testing. 0 is not tested at all.

Almost all variables had some slight change in variable values over the many years we are collecting and observing. So in our code, it is important to note both values for indicating missing data.