hw4questions

Kate Huebner

10/20/2017

To access the documentation for the FARS dataset, go to this [link](https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812447) To access the Li and Brown paper, go to this [link](http://aje.oxfordjournals.org/content/early/2014/01/27/aje.kwt327.full.pdf+html) #Question 1 ##In the in-course exercises, we have been analyzing data with accident as the observation unit. This study uses a different observation unit. What is the unit of observation in the Brady and Li study? When you download the FARS data for a year, you get a zipped folder with several different datasets. Which of the FARS datasets provides information at this observation level (and so will be the one you want to use for this analysis)? Now, the unit of observation is person or motorist, instead of accident. Their objective was to evaluate the time trends in alcohol and nonalcohol drugs detected in drivers who died within 1 hour of a motor vehicle crash in 6 states in the United States that routinely test such drivers for drugs. Thus, the 'person\_year data file' is used for thisanalysis.

# Question 2

## This study only analyzes a subset of the available FARS data. Enumerate all of the constraints that are used by the study to create the subset of data they use in their study (e.g., years, states, person type, injury type). Go through the FARS documentation and provide the variable names for the variables you will use to create filter statements in R to limit the data to this subset. Provide the values that you will want to keep from each variable.

The variables used in the analysis include st\_case, veh\_no, per\_no, state, per\_typ, lag\_hrs, lag\_mins, inj\_sev, age, alc\_res, contains("drugres"), and sex.

* st-case
* unique identifier for the state case
* veh\_no +unique identifier for the cehicle identification
* per\_no
* unique identifer for the drver
* state
* the state number

# Question 3

## The study gives results stratified by age category, year, year category (e.g., 1999–2002), alcohol level, non-alcohol drug category, and sex. For each of these stratifications, give the variable name for the FARS variable you could use to make the split (i.e., the column name you would use in a group\_by statement to create summaries within each category or the column name you would mutate to generate a categorical variable). Describe how each of these variables are coded in the data. Are there any values for missing data that you’ll need to mutate to NA values in R? Are there any cases where coding has changed over the study period?