

Case Study 3: Harvard School of Public Health College Alcohol Study

Background

The data contain four sets of survey responses from undergraduate students at 119 schools in four waves (in 1993, 1997, 1999 and 2001, respectively). The data were collected in the Harvard School of Public Health College Alcohol Study (CAS), a multi-round survey that interviewed students in four-year colleges, examined key issues in alcohol abuse and other high risk behaviors among college students, including the relationship of state alcohol control measures and college policies to alcohol use and the role of fraternities and sororities, easy access to alcohol, and low alcohol prices.

The survey collected information on students' use of alcohol, tobacco, and illicit drugs, views on campus alcohol policies and student alcohol use, reasons for drinking alcohol and reasons for not drinking or limiting drinking, and personal difficulties caused by drinking problems (e.g., missed classes, injury, and trouble with police). Additional topics covered by the survey include overall health status, daily activities, satisfaction with education being received, grade-point average, living arrangements, social life, sexual activity, use of condoms during sexual intercourse, date rape, drunk driving, and attendance in meetings of Alcoholics Anonymous, Alanon, Adult Children of Alcoholics, and Narcotics Anonymous. Background variables include age, height, weight, sex, marital status, religion, mother's and father's education, mother's and father's drinking habits, race, and Hispanic origin.

Goal

As suggested by the study title, one of the original and primary goals is to investigate factors associated with college students' drinking habits and the effects of those relevant factors. And yet, given the wide range of topics covered by the large-scale survey, the data can be used to answer other sociological/economical/psychological questions.

You may choose to answer one of the following scientific questions in your analysis **OR study a problem that you believe is interesting.**

- What are the trends of drinking, smoking or drug use among college students during the surveyed time period?
- What is the association between drinking and smoking or drug use? (e.g., is marijuana a complement or substitute to alcohol for college students?)
- What are the effects of high school drinking history, residence, race and other demographic factors on heavy drinking?
- How do students' social activities/connections (e.g., being a fraternity member, having friends who drink a lot) affect their drinking behavior?
- How does drinking/smoking/drug use affect academic performance?

Data information

The data from each survey year are included in the zip file named “Harvard_CAS_yyyy.zip” with “yyyy” representing the year. Within the zip file, you can find a folder named “DS0001”, and inside that folder there are 3 files you will need:

- ******-Codebook.pdf**: a very detailed data codebook, with variable names and explanations, as well as copies of the original surveys for your reference.
- ******-Data.txt**: the dataset; each line is a record with the variables written consecutively in an aligned manner, and blank space or a decimal point (“.”) stands for a missing value (please refer to the later pages of the codebook for how missing data were coded in different survey items).
- ******-Record layout.txt**: a table that explains the variable layout in the data file; e.g., if a variable is laid out from column 3 to 5, then the value of that variable runs from the 3rd spot to the 5th spot in a record.

Due to confidentiality reasons, the student identifier numbers (“STUDY_ID”) and their college ID numbers (“COLL_ID”) are unavailable (and thus left blank in the data file). Consequently, you won’t be able to identify individual students, or (more importantly) tell which students belong to the same school.