1) Define the sampled population and a sample. What is the difference between the two?

2) For each of the following exercises (taken from Lohr Ch 1), identify the following sample survey elements: target population, sampling frame, sampling unit, observation unit, sampled population

Also discuss briefly any problems with the sampling scheme and potential sources of nonsampling error. Remember that the sampling frame can be an actual, physical list, or merely hypothetical (think of a list of housing using versus the implicit sampling frame if we drew our sample by randomly dialing phone numbers).

a) The article “What Readers Say about Marijuana” (Parade, July 31, 1994, p. 16) reported “More than 75% of the readers who took part in an informal PARADE telephone poll say marijuana should be as legal as alcoholic beverages.” The telephone poll was announced on page 5 of the June 12 issue; readers were instructed to “Call 1-900-773-1200, at 75 cents a call, if you would like to answer the following questions. Use touch-tone phones only. To participate, call between 8 a.m. EDT [Eastern Daylight Time] on Saturday, June 11, and midnight EDT on Wednesday, June 15.”

b) Amazon books (www.amazon.com) summarizes reader reviews of the books it sells. Persons who want to review a book can submit a review online; Amazon then reports the average rating from all reader reviews on its website.

c) Potential jurors in some jurisdictions are chosen from a list of county residents who are registered voters or licensed drivers over age 18. In the fourth quarter of 1994, 100,300 jury summons were mailed to Maricopa County, Arizona, residents. Approximately 23,000 of those were returned from the post office as undeliverable. Approximately 7000 persons were unqualified for service because they were not citizens, were under 18, were convicted felons, or other reason that disqualified them from serving on a jury. An additional 22,000 were excused from jury service because of illness, financial hardship, military service, or other acceptable reason. The final sample consists of persons who appear for jury duty; some unexcused jurors fail to appear.

d) The December 2003 issue of PC World reported the results from a survey of over 32,000 subscribers asking about reliability and service for personal computers and other electronic equipment. The magazine “invited subscribers to take the Web-based survey from April 1 through June 30, 2003” and received 32,051 responses. Survey respondents were entered in a drawing to win prizes. They reported that 46% of desktop PCs had at least one significant malfunction.

3) Telling stories with data. On our first day someone mentioned that they wanted to “learn how to tell stories with data”. The Upshot’s “Nonwork” series is a great example of this. In class we read an introduction to the series and a summary of the survey they commissioned. Someone (correctly) pointed out that this survey didn’t seem large enough to support all the analyses that the authors intended to do.

Read the first full-length article in the NYT nonwork series (about nonemployed 25-54 year old men), entitled The Vanishing Male Worker: How America Fell Behind (available at http://nyti.ms/1urjh9l). Complete the following:

a) List the original sources of (quantitative) data used in the article. In addition to listing any primary data sources cited in the main text, you’ll need to do the following:
i) Open the interactive “Where Men Aren’t Working” map. Identify the source of the data being mapped and include it in your list.

ii) Click through to the American Enterprise Institute report. Identify the source of their data and include it in your list.

iii) Click through to the Brookings Institution report. Identify the source of their data and include it in your list.

iv) Read (“How Nonemployed Americans Spend Their Weekdays: Men vs. Women”) [http://nyti.ms/1xy8rEJ] and identify the source of the data used.

For parts ii-iii, it isn’t necessary to read the Brookings or AEI reports in depth.

b) For either the survey you identified in part i) or part ii), write a brief paragraph including

   i) The agency that collects the data, and for what general purpose(s)

   ii) The target population

   iii) At least one population parameter/estimand of interest

4) Generate 2 topics of interest to you for a sample survey. It could be a survey of people, things, or records. Indicate the population, one or more interesting estimands, and a possible sampling frame.