Homework 1 Solutions
303 Spring 2003
Steve Fienberg

Thursday, January 23

Part I
For each of the following surveys, describe the target population, sampling frame, sampling unit and observation unit. Discuss any possible sources of selection bias or inaccuracy of responses

- Lohr, Chapter 1 Exercise 1
  **Target Population:** All of the readers of PARADE magazine.
  **Sampling Frame:** The frame is difficult to determine. The possible respondents would be anyone who saw the magazine (or ad) on the day it appeared, however there is no list for this population.
  **Sampling Unit:** A reader who received the magazine on June 12.
  **Observation Unit:** A Reader who received the magazine on June 12 and actively decided to pay to call in and participate into the survey.
  **Selection bias/inaccuracy of response:** It was a voluntary call-in which people had to pay to participate in. This means only people who have very strong views and don’t mind paying to express them will call in. In addition, the survey was distributed in the middle of June, when children are getting out of school. It could be that many families who subscribe to the magazine that week were on vacation and didn’t have an opportunity to respond.

- Lohr, Chapter 1 Exercise 2
  **Target Population:** All of the mutual funds traded.
  **Sampling Frame:** All the mutual funds listed in the mutual fund pages of the newspaper on the day she looked.
  **Sampling Unit:** A mutual fund.
  **Observation Unit:** A mutual fund.
  **Selection bias/inaccuracy of response:** This depends on how complete the listing of mutual funds in the student’s local newspaper. Does she want mutual funds traded on all stock exchanges, or just on the NYSE or NASDAQ? Did she randomly pick the first fund out of the first 10, and then every 10 after that, or did she just take funds 10, 20, 30? There may be a zero probability of being selected depending on her scheme.

- Lohr, Chapter 1 Exercise 4
  **Target Population:** All the architects (or architect firms?) in a given city.
  **Sampling Frame:** Architects appearing in the phone book.
  **Sampling Unit:** Architecture firms appearing in the phone book.
  **Observation Unit:** An individual at the architecture firm/office.
Selection bias/inaccuracy of response: Going in alphabetical order from the phone book may cause some bias. For example, suppose that larger firms all have similar names, and smaller firms (or individuals) try to have different names. Then by ordering according to the names in the phone book, it is possible to get more of larger (or smaller) firms based on the names.

- Lohr, Chapter 1 Exercise 5
  Target Population: All books in the librarian’s library.
  Sampling Frame: A listing of locations in the library.
  Sampling Unit: A location.
  Observation Unit: An individual book at the location.
  Selection bias/inaccuracy of response: This is a good strategy based on random sampling. There may be bias depending on how books that are checked out are handled. If the librarian waits to see the condition of the book when it is returned, then the scheme is fine. If not, there may be bias against books that are often checked out (and possibly in more need of binding).

- Lohr, Chapter 1 Exercise 13
  Target Population: Any potential user of the product.
  Sampling Frame: It is not entirely clear. If they requested a group of self-identified readers of romance fiction and then sampled from that list, then the frame would be the list. It all depends on how they obtained the self-identified readers.
  Sampling Unit: An individual self-identified as a reader of romance fiction.
  Observation Unit: An individual self-identified as a reader of romance fiction.
  Selection bias/inaccuracy of response: This may be biased for a few reasons. First the frame was self-identified readers of romance fiction. They only sampled women. How were the three cities chosen? There may be bias in the cities sampled.

- Lohr, Chapter 1 Exercise 15
  Target Population: People who have HMO insurance coverage.
  Sampling Frame: Consumer Reports subscribers.
  Sampling Unit: A household (most magazine subscriptions and insurance coverage is on a household level).
  Observation Unit: The individual in the household most likely to fill out the survey (the subscriber, or the individual in charge of the health care).
  Selection bias/inaccuracy of response: There is some bias, in addition to those mentioned by the questionnaire. Only people with strong opinions tend to respond to surveys in the mail. Would only people who have HMO’s respond to the survey, or would someone without the HMO also respond?

Do you think this survey provides valuable information for comparing health plans? If you were selecting an HMO for yourself, which information would you rather have: results from this survey or results from customer-satisfaction surveys conducted by the individual HMO’s? I think it does provide interesting information, however I believe it would be biased towards negative experiences (usually people who have strong opinions have negative opinions). The advantage to this survey also is that it is done by an independent party. However, the results from a survey from an HMO may be more accurate (i.e. less biased). This is because more people may be more willing to give feedback to an HMO if they think it will result in the HMO improving service. However, if the HMO does a survey, it may be biased because they have a personal interest in looking good.