Exercise 3: Data Collection

Using the lectures, textbook and other readings as resources, answer the following questions as thoroughly and completely as possible:

**Question 1.** Identify which mode (choosing from: telephone, face to face, or mail) would be most desirable for use in a household survey, if each of the different criteria listed below were of HIGHEST priority?  
a. The survey is completed quickly

If the timeliness of survey completion was the main priority, telephone interviews would be the most desirable form of data collection as a small team of staff can complete a large number of interviews in a short time interval (Lindgren 2025).   
b. The lowest cost for a given sample size

If cost was the biggest priority for our survey, mailing would be the best option because mailing address sample frames are easily available and cheaper than accessing telephone sample frames (Lindgren 2025).   
c. The highest response rate

Face-to-face interviews are the best choice if we are prioritizing a high response rate because the presence of a trained interviewer gives a sense of credibility and a source of support for potential respondents that increase the likelihood that they will complete the survey at hand (Groves et al. 2009).

**Question 2.** Which of the three (choosing from: telephone, face to face, or mail) methods allow the researcher to most easily distinguish ineligible units?

Compared to telephone or mailing methods, face-to-face interviews are optimal for easily distinguishing ineligible units from nonrespondents (Groves et al. 2009). If sample elements do not respond to a telephone call or send back a paper questionnaire, this may either indicate ineligibility (e.g. a bad address, nonworking number) or nonrespondent units in that someone may have saw the telephone call or read the contents of a survey and chose to not respond (Groves et al. 2009). In contrast, trained interviewers can typically identify ineligible units in the field through simple observation (Groves et al. 2009).

**Question 3.** Briefly describe the challenges that cell phone only users pose for the following sources of error in a telephone survey:  
a) Coverage error

1. Overcoverage

Ineligible units may be introduced if the sampling frame contains cell phone numbers from international residents, institutionalized individuals or minors that are not included in the target population.

1. Undercoverage

For telephone surveys that use dated sample frames with only landline phone numbers, cell phone only users may be entirely excluded from sampling. This may pose a problem when interpreting the results of the survey as outcomes may differ systematically between populations with and without landline numbers.

b) Nonresponse error

The reasons for nonresponse from cell phone only users may differ from that seen in those with landline/household numbers in ways that researchers are unable to distinguish (Groves et al. 2009). Cell phone only users may have a higher level of caution when answering unknown phone numbers, and there is also a possibility that cell phone users may answer the phone be at a place where it is inconvenient to complete the survey. This may lead to nonresponse error if there are meaningful differences in the distribution of values between respondents and nonrespondents. For example, if the survey was on employment rates and employed cell phone users were less likely to answer the phone at work, this would result in biased estimates of employment.

c) Measurement error

Cell phone users may face unique challenges during survey administration that may impact measurement error. If respondents are multitasking or distracted by something on their phone while answering survey questions, their observed responses may not be accurate to the true value of the item.

**Question 4.** Shown below is the distribution of final case results for a random digit dialed (RDD) telephone survey conducted on a sample of 2127 randomly generated telephone numbers covering the contiguous 48 United States plus the District of Columbia. The target population consists of households in the 48 states and the District of Columbia. The topic of the survey is the household’s recycling activity: availability of recycling pickups in the household city/town, and the households use thereof. All phone numbers were dialed a maximum of 20 times, covering weekdays and weekends, the time and evenings, until one of these final results was achieved.  
Completed interviews 614  
Refusals 224  
Answering machine with residential message on every call 180  
Never answered every call 302  
Eligible household contacted but no interview for other than refusal reasons 127  
Business/non-residences 194  
Non-working numbers 486  
Total 2127

Compute the response rate in different ways:  
a. Assume all unanswered numbers are eligible

I referenced the AAPOR’s (2016) document of standard definitions in calculating the response rate.

R = 614/(614+224+180+127+302) = 42.5%

Assuming all unanswered numbers are eligible, the response rate for this survey is 42.5%.  
b. Assume all unanswered numbers are ineligible

R = 614/(614+224+302+127) = 53.6%

Assuming all unanswered numbers are ineligible, the response rate for this survey is 53.6%.

c. Estimate the eligibility for the unanswered numbers, called ‘e’ (Assume that the same percentage of unanswered numbers are ineligible as we found for those we could determine)

e = (614+224+180+127)/ (614+224+180+127+194+486) = 0.627

Assuming that the percentage of eligible unanswered numbers is equal to the current eligibility rate calculated by dividing the total of known eligible cases by the total of known eligible and noneligible cases, the proportion of unanswered numbers that are eligible is approximately 62.7%.   
d. Calculate the response rate using the value for e that you determined in part c

Using this value for e, the new estimate for the response rate comes out to 46%.

**References**

Groves, Robert M., Fowler, Floyd J., Jr., Couper, Mick P., Lepkowski, James M., Singer, Eleanor, and Tourangeau, Roger. 2009. *Survey Methodology*. Hoboken: John Wiley & Sons, Incorporated. ProQuest Ebook Central.

Lindgren, E. (2025). *Advantages and Disadvantages of the Four Modes* [Video]. Panopto. https://jh.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=9a93e1c1-e127-4c22-810a-b27d00fefc8c&pid=a54e0248-2a08-4496-ab1a-b281011e1d38&start=4.14449

THE AMERICAN ASSOCIATION FOR PUBLIC OPINION RESEARCH. (2016). Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. In *THE AMERICAN ASSOCIATION FOR PUBLIC OPINION RESEARCH*. https://aapor.org/wp-content/uploads/2022/11/Standard-Definitions20169theditionfinal.pdf