

**Weekly Homework I**  
Survey Methodology I  
Due date: **December 14, 2022**

1. A recent newspaper article reported that “sales of hand-held digital devices (e.g., tablets) are up by nearly 10% in the last quarter, while sales of laptops and desktop PCs have remained stagnant.” This report was based on the results of an online survey in which 9.8% of the more than 126,000 respondents said that they had “purchased a hand-held digital device between January 1 and April 30 of this year.”

Emails soliciting participation in this survey were sent to individuals using an email address frame from the 5 largest commercial Internet Service Providers (ISP) in the U.S. Data collection took place over 6 weeks beginning May 1, 2012. The overall response rate achieved in this survey was 13 percent.

Assume that the authors of this study wanted to infer something about the expected purchases of U.S. adults (18 years old +).

- a. What is the target population? What is the population in the sample frame?

*The target population is US adults, 18 years older or over. The sample frame is clients of the five largest commercial ISP, assuming that ISPs have emails of all their clients.*

- b. Briefly discuss how the design of this survey might affect the following sources of error (2 – 4 sentences each).

- Coverage error (specify the type of coverage error you are concerned with)

*The sample frame doesn't include all elements of the target population. People with other internet providers or who don't have internet access or email are excluded from the study, and could have different behavior on using technology devices.*

- Nonresponse error

*The response rate is quite low (13%). People who responded to the survey might differ from those who did not respond (e.g., younger and more involved in the use of technology). This might bias the estimation of purchases of hand-held devices.*

- Measurement error

*It's possible respondents don't remember exactly when they bought a device. Memory problems could overestimate purchases (e.g., including purchases that occurred a year ago, not the period of interest).*

- c. Without changing the duration or the mode of this survey (i.e., computer-assisted, self-administration), what could be done to reduce the errors you outlined in 1b? For each source of error, suggest one change that could be made to reduce this error component, making sure to justify your answer based on readings and lecture material (1– 2 sentences each).

*Regarding coverage error, I would suggest extending the sample frame to include other directories of emails (small ISP providers, institutional providers). For non-response error, I would provide better incentives (e.g., prizes) to respondents to increase participation rates. In regard to measurement error, I would change the time reference of the question: Have you bought some hand-held devices (e.g., tables, kindles, etc.) during the last four months?*

- d. To lower the cost of this survey in the future, researchers are considering cutting the sample in half, using an email address frame from only the 2 largest ISPs. What effect (if any) will these changes have on sampling error and coverage error (1 – 3 sentences each)?

*The precision of purchase estimates will decrease (more sampling error), assuming the same variance, statistician confidence, and we know the probability of selection of units. Coverage error, on the other hand, might increase because email addresses from the two largest ISP can have a different profile of users (e.g., older, from big cities or specific areas of the US).*

2. For each of the following design decisions, identify which error source might be affected. Each design decision can affect at least two different error sources. Write short (2-4 sentences) answers to each point.

- a) The decision to use repeated calls persuading reluctant respondents in a customer satisfaction survey for a household product manufacturer.

*Using repeated calls might reduce non-response error as people who respond to a customer satisfaction survey could be happier with the manufacturer. However, it could affect the quality of measurement: respondents might declare less satisfaction due to getting repeated calls.*

- b) The decision to increase the number of questions about assets and income in a survey of income dynamics, resulting in a lengthening of the interview.

*While increasing the number of questions might improve the validity of income dynamics measurement, it could also reduce the quality of information due to respondent fatigue. Longer questionnaires would also increase item non-response.*

- c) The decision to include prisons and hospitals in the sampling frame for a study of consumer expenditures.

*Assuming that people from prisons and hospitals are not active consumers, to include them corresponds to a specification error (units not suitable to the goals of the study), and coverage error (inclusion of ineligible units).*

- d) The decision to change from a face-to-face interview design to a mailed questionnaire mode in a household survey of illegal drug usage.

*Using a mailed questionnaire might improve the report of illegal drug use (lower measurement error due to the reduction of social desirability bias), but it can also increase non-response error (e.g., problematic users will probably not respond to a mailed questionnaire).*

3. Five years after the last census, you mount a household survey using a telephone number frame. If a selected telephone number is a household number, interviewers ask to speak to the person most knowledgeable about the health of the household members. After the survey is over, someone suggests evaluating your survey by comparing the demographic distributions (i.e., age, sex, race/ethnicity) of your “most knowledgeable” health informant to the demographic distributions of adults from the last census. Comment on the wisdom of this suggestion.

*It is a good idea to compare the distribution of sociodemographic variables that comes from a telephone survey with census data, assuming that we are contrasting the same geographic areas and demographic groups. However, we would need to compare the distribution of all household members to the census distribution, not only the most knowledgeable, as we know they would probably be women or older. Comparing distributions from all household members would give us a better idea of the potential demographic biases of our telephone survey. This won't resolve measurement issues related to the health report.*