# Exercise 1: Populations, Frames, and Coverage Error

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

Question 1. Using some of the issues discussed in the reading, discuss some of the errors  
made in the spectacular failure of the 1936 Literary Digest poll? (1-2 paragraphs)

One of the most infamous failures in survey methodology was the 1936 Literary Digest poll, which predicted a landslide victory for Republican Alfred Landon in the presidential election based on responses from millions of Americans. However, the actual winner was the Democratic candidate, Franklin D. Roosevelt, who carried 46 out of 48 states. This catastrophic inaccuracy resulted not from a single mistake but from multiple methodological errors throughout the survey process.

One key issue was coverage error—systematic differences between the sample frame and the overall target population. The survey relied on Literary Digest subscribers, a group that was likely quite different from the general electorate in several ways, including political ideology and socioeconomic status. During the Great Depression, many Americans could not afford basic necessities, let alone magazine subscriptions. As a result, the survey disproportionately represented wealthier individuals, who were more likely to support Landon, while underrepresenting lower-income voters, who overwhelmingly backed Roosevelt. This imbalance in the sampling frame led to coverage bias in the poll’s estimates of candidate preferences.

Beyond coverage error, other sources of bias emerged through sampling and nonresponse errors. Of the 10 million surveys mailed out, only 2.4 million were returned (Lusinchi, 2016). Those who responded were likely to be more politically engaged and, in this case, more Republican-leaning. Because the poll used a non-probability sample, this bias in the estimates can be attributed to both sampling error and nonresponse error, as the political preferences of non-respondents systematically differed from those who participated. Ultimately, these flaws combined to create an inaccurate and misleading projection of the election outcome.

Question 2. For each of the following surveys, identify the target population, the sample frame, and one possible coverage issue (this could be an over or an under coverage issue). Then for each,  
propose one change to the sample frame that you think would improve coverage and reduce  
error. Support your proposal with evidence from the reading and lecture.

**1) NSDUH**  
A) Target Population:

The target population of the NSDUH includes all noninstitutionalized US citizens aged 12 and older (Groves et al. 2009). This excludes those in this age group who are living in institutions such as prisons, hospitals or military barracks at the time of the survey.  
B) Sample Frame:

The sample frame for the NSDUH is US households, which requires multiple stages from selecting area units like counties or blocks, designating addresses within a given area unit and then obtaining listings of persons within each household (Groves et al. 2009).

C) Coverage Error:

One source of coverage error in the NSDUH is the underrepresentation of individuals who are often excluded from a household-based area frame due to unique housing conditions. For example, the book notes that young men of color are particularly at risk of being omitted from sampling frames because they may move frequently between multiple residences, such as their own homes and those of friends or relatives (Groves et al., 2009). Similarly, transient or mobile populations, including individuals living in mobile homes or migrant workers, are also likely to be excluded from the sampling frame. As a result, coverage bias may occur if these excluded subgroups differ systematically from those included in the sample on key survey measures, such as recreational drug use or prescription medication use (Lindgren, 2025). If young men of color who are omitted from the survey have higher rates of recreational drug use than those captured in the sampling frame, the resulting statistics would underestimate overall drug use rates for the entire target population.

D) Proposed Frame change:

I believe multiplicity sampling, which leverages networks within the sample frame to recruit peers, could help improve coverage and reduce error in the NSDUH. This approach addresses the underrepresentation of young men of color and transient populations, who are often missed in traditional household-based sampling frames. In this method, a network member within the sample frame can provide a list of peers, who can then be subjected to specific weighting procedures to ensure they have an equal chance of being selected for the survey sample (Groves et al., 2009).

**2) NAEP**  
A) Target Population:

The target population of the national NAEP is schoolchildren in grades 4, 8, and 12 residing in the 50 states and the District of Columbia. The target population of the state NAEP is schoolchildren in grades 4 and 8 (Groves et al., 2009).

B) Sample Frame:

Like the NSDUH, the sample frame for the NAEP also consists of several stages; a list of schools is compiled from state and national education databases, and within the schools that are randomly selected, a sample of schoolchildren is taken from listings of all students in grades 4, 8, and 12 (Groves et al., 2009).

C) Coverage Error:

Because the sample frame for the NAEP relies on databases of public and private educational institutions, it excludes homeschooled students and highly transient students such as those in foster care or homeless situations that may not be in school long enough to be an eligible element of the sample frame. Homeschooled and highly mobile student populations may have substantial baseline differences when compared to students registered in traditional school settings included in the sample frame, possibly in ways that are relevant to the outcome of educational achievement (Lindgren, 2015). For example, highly mobile students may also demonstrate a greater prevalence of mental health/family-related challenges which may affect patterns in academic performance in ways that are not accounted for by estimates taken from the current sample frame.

D) Proposed Frame change:

A multiple frame design would help increase coverage for both highlighted subgroups of the student population currently omitted from the sample frame of the NAEP. In addition to the main sample frame of public and private schools in state/federal education databases, the survey could also use listings of schoolchildren provided by county or state level homeschool or Charter school organizations, while also ensuring that elements in both frames are not duplicated (Groves et al., 2009). Similarly, utilizing listings from community youth organizations or foster care institutions would help expand the NAEP sample frame to be more representative of the actual target population.

Question 3. Answer the questions about the following survey. (Adapted from Survey Methodology  
ch.2 ex.1) A recent newspaper article reported that “sales of handheld digital devices 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 they “purchased a handheld 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 five largest commercial Internet service providers in the US. Data collection took  
place over a 6- week period beginning May 1, 2002. The overall response rate achieved was  
13%.

Assume that the authors of the study wanted to infer something about the purchases of U.S.  
Adults (18 years old and older).  
1. What is the target population?

The target population is US adults aged 18 or older in 2002.

2. What is the population in the sample frame?

The population in the sample frame is elements with email addresses from the five largest commercial internet service providers in the US.

3. Is this a probability sample?

No, it is not because not every element of the sample frame has a known, nonzero chance of being selected for the survey sample (Levy and Lemeshow, 2008). Participation was voluntary, and those in the sample frame who chose to complete the survey may differ in non-random ways compared to those who did not.

Discuss how the design of this survey might affect the following:  
4. Coverage error

The sample frame does not include US adults who did not have an email address (which could be a significant proportion given that this survey was conducted in 2002). Specifically, older adults, low-income individuals and those in rural areas were less likely to have email addresses or internet access in 2002. These omitted population subgroups can additionally be expected to engage less with the handheld digital device market than elements included in the sample frame who are also likely even more tech savvy than the population average, thus the percentage of US adults who bought a handheld digital device in the last quarter may be an overestimate of the actual population parameter.

5. Nonresponse error

With a high nonresponse rate of 87%, it is possible that those in the sampling frame who chose to respond differ systematically to those who did not opt in. If people who did buy digital devices recently were more likely to respond because they have had positive experiences or are excited about their technology, then the estimate of consumer behavior presented is likely an overstatement as those who did not partake in the survey were also less likely to have bought digital devices recently.

6. Measurement error

Without additional context or examples, respondents may have interpreted the question regarding the purchase of a handheld digital device to include more or less items than what was originally intended by survey researchers. Additionally, because the item is asking respondents about their purchases in the last quarter, there is always the possibility that respondents misremembered purchase dates or exaggerated their buying habits, so that the observed survey responses of participants are different to the true value of the measurement.

**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.

Levy, P. and Lemeshow, S. (2008). Sampling of Populations: Methods and Applications, 4th ed, NY: Wiley

Lindgren, E. (2025). *Coverage Error* [Video]. Panopto. https://jh.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=23a75e46-91f6-4e3f-b95d-b25601713fc4&pid=49b58183-17ce-4a4d-ad92-b26800fe96b2&start=789.74892

Lusinchi, D. (2016). “President” Landon and the 1936 Literary Digest Poll: Were Automobile and Telephone Owners to Blame? *Social Science History*, *36*(1), 23–54. https://doi.org/10.1215/01455532-1461650