

# 2 Elements of the Sampling Problem

## 2.1 Introduction

The **design of the survey** is the method for determining who will and will not be part of the sample.

Some examples include **probability sampling** and **quota sampling**.

What types of *errors* occur when sampling, and how can these errors be avoided?

How should a questionnaire be designed, and what are the steps for performing a valid survey?

## 2.2 Technical Terms

A **population** or **target population** is a collection of elements about which we wish to make an inference.

The **sampling unit** is what we actually sample.

**Example:** James Wright (1988, *Social Problems* 35, pp. 182-192) estimates that 33% of all homeless people are mentally ill, by sampling homeless persons who received medical attention from one of the clinics in the Health Care for the Homeless project.

A **(sampling) frame** is the list of sampling units.

An **element** (also called an **observation unit**) is an object on which a measurement is taken.

A **sample** is a collection of sampling units drawn from a frame or frames.

Can Wright's estimate be fully trusted?

**Example:** Ann Landers (1976, *Good Housekeeping* 182 (June), pp. 100-101, 215-216, 223-224) asked readers of her column to respond to this question: "If you had it to do over again, would you have children?" About 70% of the readers who responded said no. She received over 10,000 responses, 80% of those from women.

What are the population, sampling frame, sampling unit, and element?

Can we trust her results?

## 2.3 How to Select the Sample: The Design of the Sample Survey

*Notation:* The unknown (fixed, non-random, population) parameter (to be estimated) is  $\theta$ , and the (random, based on data) estimator of  $\theta$  is denoted by  $\hat{\theta}$ .

The **error of estimation** is  $|\hat{\theta} - \theta|$ , which we want to minimize.

Select error bound  $B > 0$  such that  $|\hat{\theta} - \theta| < B$  with high probability.

Suppose the estimator  $\hat{\theta}$  is approximately normal for a large sample.

The bound  $B$  often is called the **margin of error**.

How can we force  $B$ , the margin of error, to be sufficiently small?

## PROBABILITY SAMPLING

**Probability sampling** implies that the sampling procedure depends on planned randomness.

Types of probability sampling (simple random sampling, stratified random sampling, cluster sampling, and systematic sampling) are below:

(A) **Simple random sampling** occurs when the sample size  $n$  is selected in advance, such that each group of  $n$  individuals in the sampling frame is equally likely to be selected.

Is this sampling procedure reasonable?

Suppose a medical technician needs to know your red blood cell count.

(B) **Stratified random sampling** occurs when the population is divided into groups called *strata*, such that a simple random sample is taken within each *stratum*.

For example: Poll elementary school children in Harrisonburg to inquire about whether or not they like school, or get enough food on a daily basis.

How should stratification be done?

(C) **Cluster sampling** occurs when the elements (or observation units) are aggregated into large sampling units, called *clusters*, such that some clusters are sampled, and then additional sampling occurs within the clusters themselves.

For example, suppose you want to survey Lutheran church members in Minneapolis, but you do not have a list of all church members in the city, so you cannot take a simple random sample of church members.

Does a *cluster sample* of 500 Lutherans provide as much information as a *simple random sample* of 500 Lutherans?

Why should *cluster sampling* be performed?

**Example:** Suppose one wants to estimate the average amount of time that professors at JMU spend grading homework in a specific week.

How can one perform a *simple random sample* of size  $n$ ?

How can one perform a *stratified sample*?

What advantage does *stratified sampling* have over *simple random sampling*?

How can one perform a *cluster sample*?

For this example, is *cluster sampling* better than *stratified sampling*?

**SYSTEMATIC SAMPLING** - Sampling is done on every  $k$ th item in a list.

Most of the time, a *systematic sample* gives results comparable to those of a *simple random sample*.

When might a *systematic sample* not be as satisfactory as a *simple random sample*?

When might a *systematic sample* improve over a *simple random sample*?

*Caution:* In systematic sampling, we still must have a sampling frame and be careful when defining the target population.

Is sampling every 20th student who enters the library a representative sample of the entire student body?

Is sampling every 10th passenger who exits an airplane a representative sample of all the persons on that flight?

**QUOTA SAMPLING** - The population is divided into different subpopulations just as in stratified random sampling, but with one important difference: Probability sampling is not used to select individuals from the subpopulation for the sample.

**Example:** Poll 100 students regarding on-campus food services. Suppose 60% of the students are female, and 40% are male.

How would quota sampling be performed?

Can the results be trusted?

**Example:** Election of 1948

The major polls were predicting that Thomas Dewey would defeat Harry Truman.

The Gallup Poll predicted 44.5% for Truman and 49.5% for Dewey.

However, in the actual election, Truman received 49.5% and Dewey received 45.1%.

What was the main source of error from these major polls?

General instructions given to interviewers might be: “Find two men and three women in your block, and make sure four are over 25 years of age and one is under 25.”

## 2.4 Sources of Errors in Surveys

Two major sources of errors in surveys are *errors of nonobservation* (e.g., caused by how respondents are selected to be in the sample) and *errors of observation* (e.g., caused by confusing questionnaire, biased-sounding interviewer, dishonest or exaggerating respondent).

### ERRORS OF NONOBSERVATION

Errors of nonobservation are caused by *sampling*, *coverage*, or *nonresponse*.

The **sampling error** is the error that results from taking a sample rather than examining the whole population.

Errors due to **coverage** occur when the sampling frame does not match up perfectly with the target population.

Errors due to **nonresponse** occur when the person selected for the survey does not participate.

**Example:** To study nutrient content of menus in boarding homes for the elderly in Washington State, a researcher (Goren et al. 1993, *Journal of Nutrition for the Elderly*) mailed surveys to all 184 licensed homes in the state, directed to the administrator and food service manager. Of those, 43 were returned by the deadline and included menus.

Can results from the above example be trusted?

**Double sampling** (see section 5.11 on pp. 153-155) and **poststratification** (see section 11.6 on pp. 366-373: Adjusting for Nonresponse) are techniques which can be used to reduce the bias caused by nonresponse.

Avoid surveys based on **volunteers**.

Following President George H. W. Bush's State of the Union Address on January 28, 1992, CBS News conducted a call-in poll.

Can the results from this poll be trusted?

The *New York Times* referred to this study as "the largest biased sample in the history of instant polling."

## ERRORS OF OBSERVATION

Errors of observation are caused by the interaction between the interviewer/questionnaire and the subject being interviewed.

### Examples of errors of observation:

⊙ People might not tell the truth.

⊙ People might not understand the questions.

Recall the news statement in 1993: “Twenty-two percent of Americans doubt that the Holocaust ever occurred.”

⊙ **Telescoping** (which occurs in the design of the National Crime Victimization Survey).

⊙ People give different answers to different interviewers.

Schuman and Converse (1971, *Public Opinion Quarterly*) interviewed black residents of Detroit.

Surveyed question: “Do you personally feel that you can trust most white people, some white people, or none at all?”

What do you think the response was, regarding “most white people”?

⊙ People may say what they think an interviewer wants to hear or what they think will impress the interviewer.

Surveyed question: “Do you agree or disagree with the following statement?”

Lenski and Leggett (1960, *American Journal of Sociology*) used the following two statements simultaneously on a survey:

“It is hardly fair to bring children into the world, the way things look for the future.”

“Children born today have a wonderful future to look forward to.”

⊙ Interviewer might misread question, record responses inaccurately, or antagonize the respondent.

⊙ Certain words have different meanings to different people.

What does the following mean to you: “Do you own a car?”

⊙ Question wording and order.

Two different surveys taken in late 1993 and early 1994 asked these questions:

“In the past few years, there have been a lot of rumors and stories about whether Elvis Presley is really dead. How do you feel about this? Do you think that there is any possibility that these rumors are true and that Elvis Presley is still alive, or don’t you think so?”

“A recent television show examined various theories about Elvis Presley’s death. Do you think it is possible that Elvis is alive or not?”

### **Error of nonobservation or error of observation?**

⊙ “Does it seem possible or does it seem impossible to you that the Nazi extermination of the Jews never happened?”

- ⊙ The person sampled says, “I’m too busy to respond to a survey. I need to get to the country club.”
- ⊙ Poll 100 JMU students to ask: “Do you approve or disapprove of the mayor of New York City?”

#### **Four main types of procedures for collecting data**

- ⊙ Personal interviews
- ⊙ Telephone interviews
- ⊙ Self-administered questionnaire
- ⊙ Direct observation

#### **Reducing errors in surveys**

- ⊙ Callbacks
- ⊙ Rewards and incentives

- ⊙ Trained interviewers

- ⊙ Data checks

Ask questions which should provide similar results.

- ⊙ Questionnaire construction

## 2.5 Designing a Questionnaire

Below are some concerns when designing a questionnaire.

### QUESTION ORDERING

- ⊙ McFarland (1981, *Public Opinion Quarterly*) conducted a survey with the *general* question, “How interested would you say you are in religion: very interested, somewhat interested, or not very interested?”

McFarland asked the *specific* question, “Did you, yourself, happen to attend church in the last seven days?”

Depending on the order of these two questions, what percentage responded “very interested in religion”?

- ⊙ Serdula et al. (1995, *American Journal of Epidemiology*) reported results from a health survey with many questions:

The respondent was asked:

- (1) to report his/her weight,  
and was asked as the next question,
- (2) “Are you trying to lose weight?”

In the same survey with a different question ordering, question #2 was asked toward the middle of the survey, and question #1 was asked at the end of the survey.

### Ordering of responses

- ⊙ In a **written** list of several choices, the choices near the top of the list are more likely to be selected than if those same choices had been placed at the end of the list.

For example, choices might include *strongly agree*, *agree*, *disagree*, or *strongly disagree*.

What can be done to remove some of this bias caused by the ordering of the responses?

- ⊙ From an **oral** list of several choices with **lengthy** choices, the respondent may tend to forget the choices and hence select the most recent choice.

### OPEN-VERSUS-CLOSED QUESTIONS

In a **closed** question, the respondent will give either a single numerical answer (such as age), or will make a selection from multiple choice options.

In an **open** question, the respondent is not prompted with the interviewer’s preselected categories.

Which is preferred, *open* or *closed*?

- ⊙ Bradburn and Sudman (1979, *Improving Interview Method and Questionnaire Design*) noted that respondents reported higher frequency of drinking alcoholic beverages when asked an *open* question than a *closed* question with categories “never” through “daily.”
  
- ⊙ The National Crime Victimization Survey, using a questionnaire in 1991, asked, “Has anyone attacked or threatened you in any of these ways: (a) With any weapon, for instance, a gun or knife, (b) With anything like a baseball bat, frying pan, scissors, or stick? ...”
  
- ⊙ *Closed* vs. *open* questions may produce different responses.

The survey by Skelly et al. (1968) on women’s attitudes toward fabrics used in clothing gave about half the sample an *open* version of the questionnaire, and the other half a *closed* version.

First question in the *open* questionnaire: “What difficulties and problems do you run into most often when buying clothes, any kind of clothes, for yourself?”

Corresponding first question in the *closed* questionnaire: “Which of these reasons best describes the difficulties and problems you run into most often when buying clothes, any kind of clothes, for yourself? Any others?”

- |                                      |                                     |
|--------------------------------------|-------------------------------------|
| 1. I am short waisted.               | 8. I have wide hips.                |
| 2. I am long waisted.                | 9. Limited styles, selections.      |
| 3. I need a short length.            | 10. I have problems with necklines. |
| 4. I need a long length.             | 11. Can’t find correct sizes.       |
| 5. I have a small waist.             | 12. Sizes don’t run true.           |
| 6. I have a large waist.             | 13. Poor workmanship.               |
| 7. Doesn’t fit around the shoulders. |                                     |

How did the responses differ for *closed* vs. *open* questionnaires?

If using a *closed* question, consider having an “other” category. Here is why!

- ⊙ In one study of sexual activity among adolescents, adolescents were asked from whom they felt the most pressure to have sex. Categories for the *closed* question were “friends of same sex,” “boyfriend/girlfriend,” “friends of opposite sex,” “TV or radio,” “don’t feel pressure,” and “other.”

## RESPONSE OPTIONS

Should or should not “no opinion” be an option?

**Example:** “Do you think the enforcement of traffic laws in our city is too strict or too lenient?”

## WORDING OF QUESTIONS

- ⊙ Make sure the questions are clearly stated.
- ⊙ Use forced-choice, rather than agree/disagree, questions.

Schuman and Presser, (1981, *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*, p. 223) asked the following questions in *separate* surveys:

*Question 1:* “Do you agree or disagree with this statement: Most men are better suited emotionally for politics than are most women.”

*Question 2:* “Would you say that most men are better suited emotionally for politics than are most women, that men and women are equally suited, or that women are better suited than men in this area?”

	Years of Schooling		
	0 – 11	12	13+
Q1: percent “agree”	57	44	39
Q2: percent “men better suited”	33	38	28

- ⊙ Avoid questions that prompt or motivate the respondent to say what you would like to hear.

Such questions are called **leading** or **loaded**.

The *Wall Street Journal* (May 17, 1994) reported the following question asked by the Gallup Organization in a survey commissioned by the American Paper Institute.

“It is estimated that disposable diapers account for less than 2 percent of the trash in today’s landfills. In contrast, beverage containers, third-class mail and yard waste are estimated to account for about 21 percent of trash in landfills. Given this, in your opinion, would it be fair to tax or ban disposable diapers?”

- ⊙ Ask only one concept in each question. Avoid **double-barreled questions**.

A member of the U.S. House of Representatives surveyed the following question to his constituents:

“Do you agree with Bill Clinton’s \$50 billion bailout of Mexico?”

## 2.6 Planning a Survey

- (1) Statement of objectives.

- (2) Target population.
- (3) The frame.
- (4) Sample design.
- (5) Method of measurement.
- (6) Measurement instrument.
- (7) Selection and training of field-workers.
- (8) The pretest.
- (9) Organization of fieldwork.
- (10) Organization of data management.
- (11) Data analysis.

## 2.7 Summary

Some useful sampling designs are *simple random sampling*, *stratified sampling*, *cluster sampling*, and *systematic sampling*.

We want the *sampling frame* to be similar to the *target population*.

Clearly worded questionnaire is needed for a good survey.