Monday, October 29, 2018

- **Warm-up**
  - **Answer the following 2 questions. Explain your reasoning.**

In a survey of public opinion concerning state aid to a particular city, every 40th person registered as a voter was interviewed, beginning with a person selected at random from among the first 40 listed. This is an example of

A) Simple random sampling
B) Systematic sampling
C) Stratified sampling
D) Single-stage cluster sampling
E) None of the above

- **Sampling & sample statistics**

To survey the opinions of the students at your high school, a researcher plans to select every twenty-fifth student entering the school in the morning. Assuming there are no absences, will this result in a simple random sample of students attending your school?

(A) Yes, because every students has the same chance of being selected.

(B) Yes, but only if there is a single entrance to the school.

(C) Yes, because the 24 out of every 25 students who are not selected will form a control group.

(D) Yes, because this is an example of systematic sampling, which is a special case of simple random sampling.

(E) No, because not every sample of the intended size has an equal chance of being selected.

Social Objective: I will participate in the class activities.

Content Objective: I will apply different types of random sampling to my class.

Language Objective: I will describe the types of random sampling clearly both verbally and in writing (in my notes).

Does a Census Make Sense?

- Why bother determining the right sample size?
- Wouldn't it be better to just include everyone and "sample" the entire population?
  - Such a special sample is called a census.

Populations and Parameters

- Models use mathematics to represent reality.
  - Parameters are the key numbers in those models.
- A parameter that is part of a model for a population is called a **population parameter**.
- We use data to estimate population parameters.
  - Any summary found from the data is a statistic.
  - The statistics that estimate population parameters are called **sample statistics**.

Notation

- We typically use Greek letters to denote parameters and Latin letters to denote statistics.

Name	Statistic	Parameter
Mean	$\bar{y}$	μ (mu, pronounced "meeoo," not "moo")
Standard deviation	s	σ (sigma)
Correlation	r	ρ (rho)
Regression coefficient	b	β (beta, pronounced "baytah"[5])
Proportion	$\hat{p}$	p (pronounced "pee"[6])

Samples

- We draw samples because we can't work with the entire population.
  - We need to be sure that the statistics we compute from the sample reflect the corresponding parameters accurately.
  - A sample that does this is said to be **representative**.

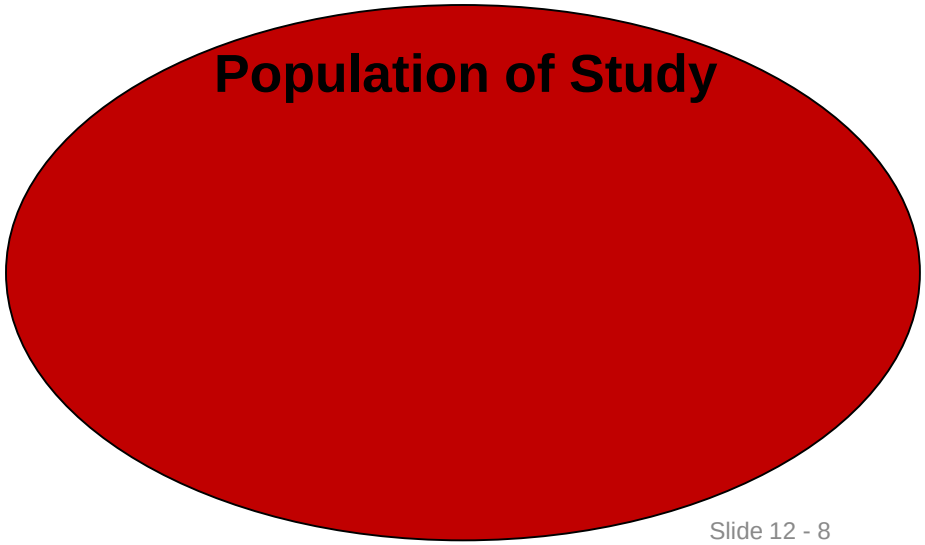
Simple Random Samples

- We will insist that every possible *sample* of the size we plan to draw has an equal chance to be selected.
  - Such samples also guarantee that each individual has an equal chance of being selected.
  - With this method each *combination* of people has an equal chance of being selected as well.
  - A sample drawn in this way is called a **Simple Random Sample (SRS).**
- An SRS is the standard against which we measure other sampling methods, and the sampling method on which the theory of working with sampled data is based.

Defining the "Who"

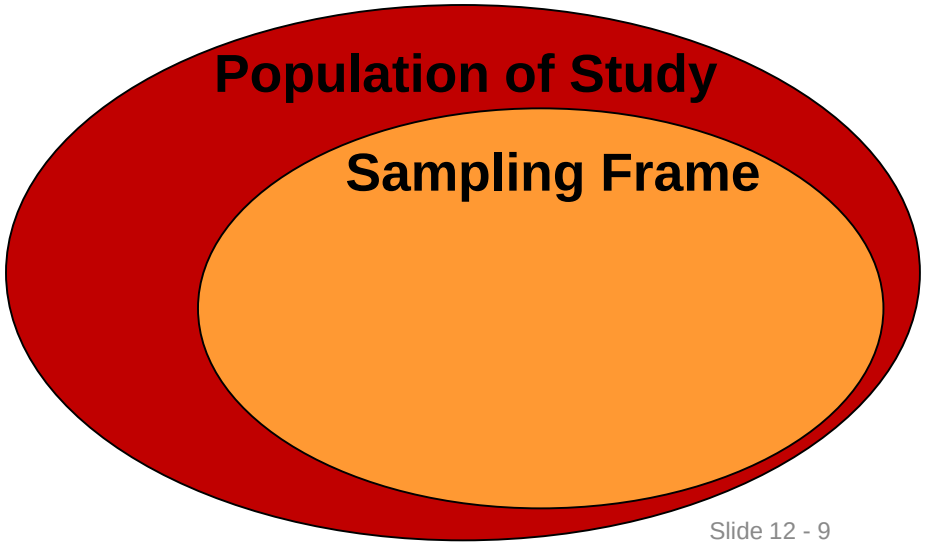
- The *Who* of a survey can refer to different groups, and the resulting ambiguity can tell you a lot about the success of a study.
- To start, think about the population of interest. Often, you'll find that this is not really a well-defined group.
  - Even if the population is clear, it may not be a practical group to study.

1st step



Defining the "Who" (cont.)

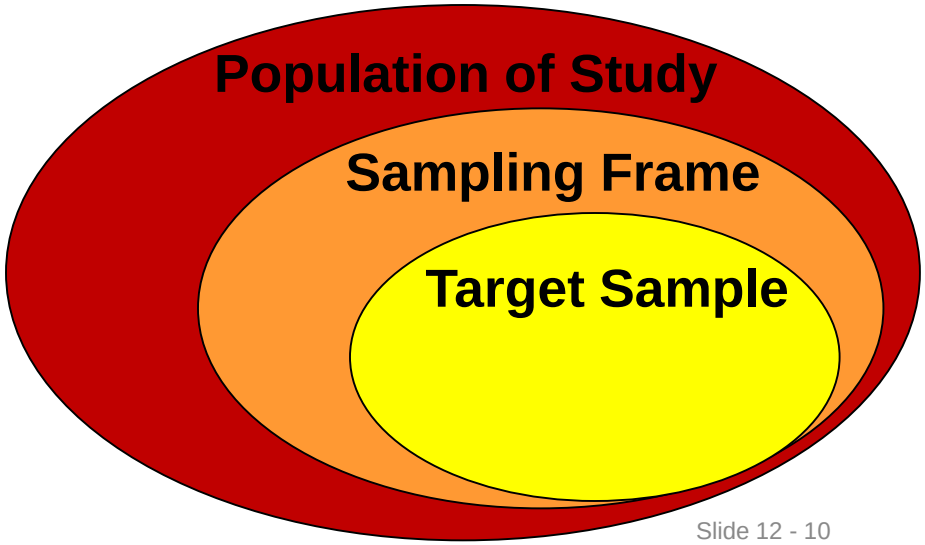
- Second, you must specify the sampling frame.
  - Usually, the sampling frame is not the group you *really* want to know about.
  - The sampling frame limits what your survey can find out.



Defining the "Who" (cont.)

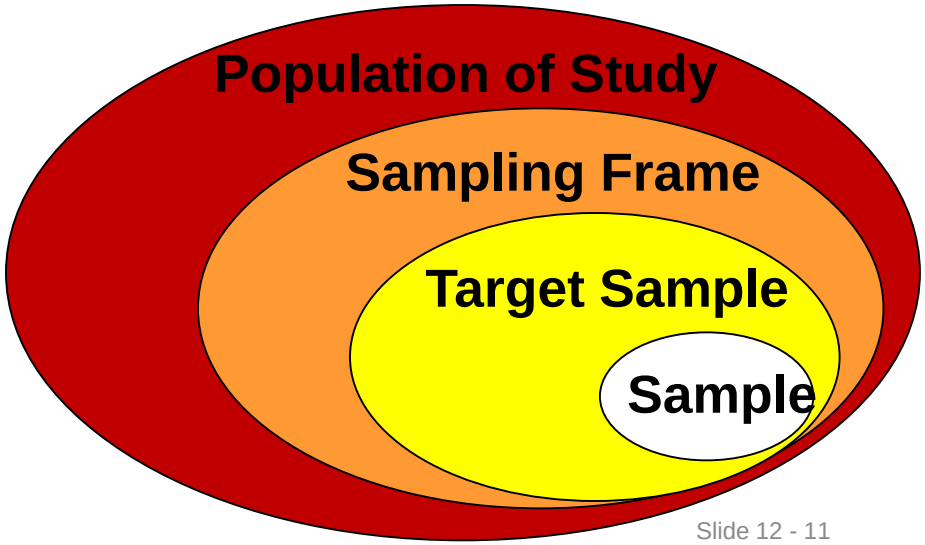
- Then there's your target sample.
  - These are the individuals for whom you *intend* to measure responses.
  - You're not likely to get responses from all of them. Nonresponse is a problem in many surveys.

NCHS
Cherry Creek



Defining the "Who" (cont.)

- Finally, there is your sample—the actual respondents.
  - These are the individuals about whom you *do* get data and can draw conclusions.
  - Unfortunately, they might not be representative of the sample, the sampling frame, or the population.



To select a sample at random, we first need to define where the sample will come from.

The **sampling frame** is a list of individuals from which the sample is drawn.

Once we have our sampling frame, the easiest way to choose an SRS is to assign a random number to each individual in the sampling frame.

- **Samples drawn at random generally differ from one another.**
  - **Each draw of random numbers selects *different* people for our sample.**
  - **These differences lead to different values for the variables we measure.**
  - **We call these sample-to-sample differences sampling variability.**

Multistage Sampling

- Sometimes we use a variety of sampling methods together.
- Sampling schemes that combine several methods are called **multistage samples**.
- Most surveys conducted by professional polling organizations use some combination of stratified and cluster sampling as well as simple random sampling.

The Valid Survey

- It isn't sufficient to just draw a sample and start asking questions. A *valid* survey yields the information we are seeking about the population we are interested in. Before you set out to survey, ask yourself:
  - What do I want to know?
  - Am I asking the right respondents?
  - Am I asking the right questions?
  - What would I do with the answers if I had them; would they address the things I want to know?

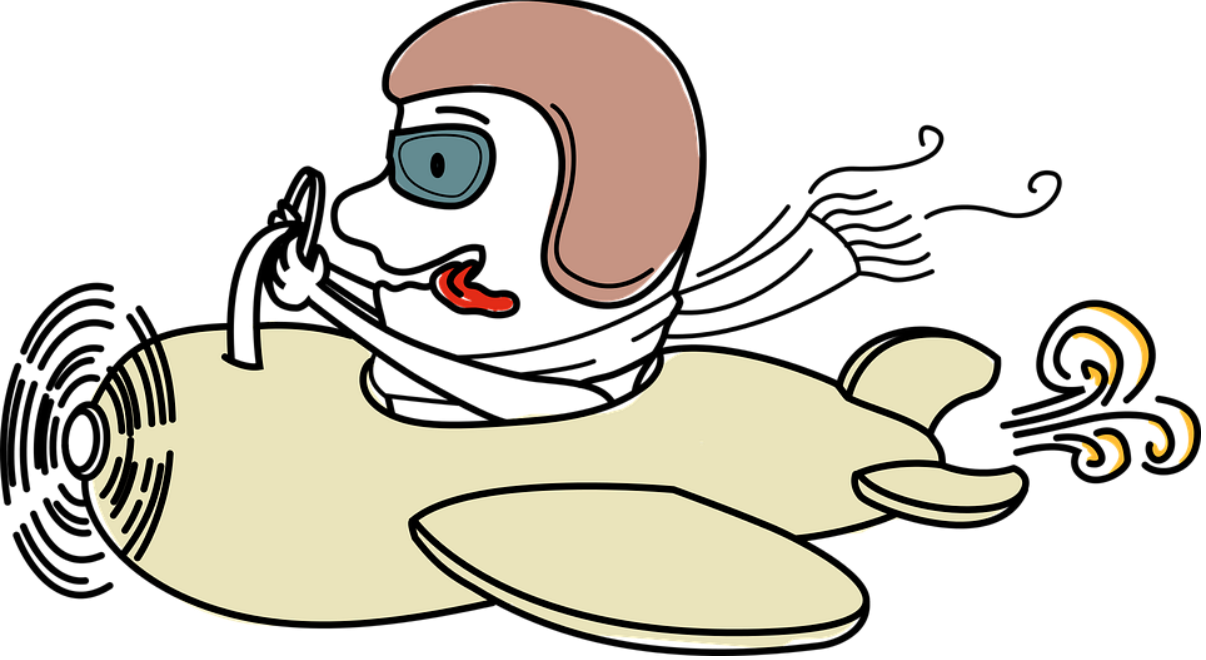
The Valid Survey (cont.)

These questions may sound obvious, but there are a number of pitfalls to avoid.

- Know what you want to know.
  - Understand what you hope to learn and from whom you hope to learn it.
    - Use the right frame.
      - Be sure you have a suitable sampling frame.
        - Tune your instrument.
          - The survey instrument itself can be the source of errors - too long yields less responses.

The Valid Survey (cont.)

- Ask specific rather than general questions.
- Ask for quantitative results when possible.
- Be careful in phrasing questions.
  - A respondent may not understand the question or may understand the question differently than the way the researcher intended it.
- Even subtle differences in phrasing can make a difference.
- Be careful in phrasing answers.
  - It's often a better idea to offer choices rather than inviting a free response.



The Valid Survey (cont.)

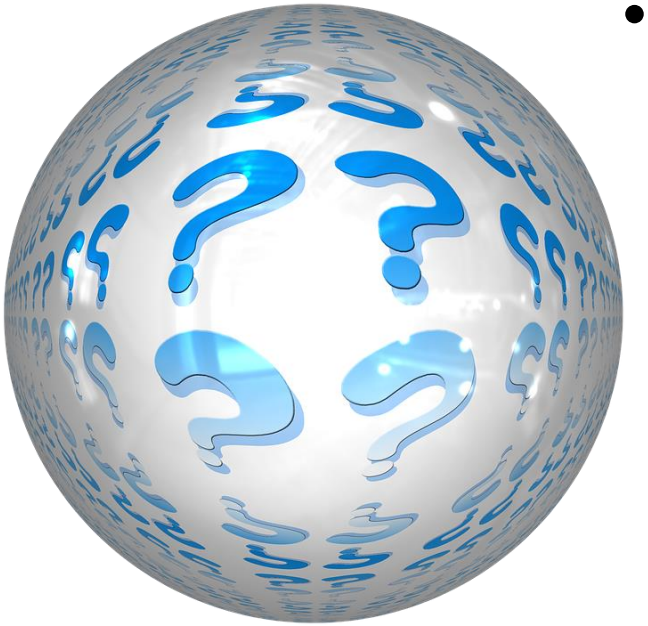
- The best way to protect a survey from unanticipated measurement errors is to perform a pilot survey.
  - A pilot is a trial run of a survey you eventually plan to give to a larger group.

What Can Go Wrong?—or, How to Sample Badly

- Sample Badly with Volunteers:
  - In a voluntary response sample, a large group of individuals is invited to respond, and all who do respond are counted.
    - Voluntary response samples are almost always biased, and so conclusions drawn from them are almost always wrong.
  - Voluntary response samples are often biased toward those with strong opinions or those who are strongly motivated.
  - Since the sample is not representative, the resulting voluntary response bias invalidates the survey.

What Can Go Wrong?—or, How to Sample Badly (cont.)

- Sample Badly, but Conveniently:
  - In convenience sampling, we simply include the individuals who are convenient.
    - Unfortunately, this group may not be representative of the population.
  - Convenience sampling is not only a problem for students or other beginning samplers.



- In fact, it is a widespread problem in the business world—the easiest people for a company to sample are its own customers.

What Can Go Wrong?—or, How to Sample Badly

- **Sample from a Bad Sampling Frame:**
  - An SRS from an incomplete sampling frame introduces bias because the individuals included may differ from the ones not in the frame.
- **Undercoverage:**
  - Many of these bad survey designs suffer from **undercoverage**, in which some portion of the population is not sampled at all or has a smaller representation in the sample than it has in the population.
  - **Undercoverage** can arise for a number of reasons, but it's always a potential source of bias.

What Else Can Go Wrong?

- Watch out for nonrespondents.
  - A common and serious potential source of bias for most surveys is **nonresponse bias.**
  - No survey succeeds in getting responses from everyone.

- The problem is that those who don't respond may differ from those who do.
- And they may differ on just the variables we care about.

What Else Can Go Wrong? (cont.)

- Don't bore respondents with surveys that go on and on and on and on…
  - Surveys that are too long are more likely to be refused, reducing the response rate and biasing *all* the results.
- Work hard to avoid influencing responses.
  - Response bias refers to anything in the survey design that influences the responses.
  - For example, the *wording* of a question can influence the responses:

1. A uniformed policeman interviews a group of 50 college freshmen. He asks each one his or her name and then if he or she has used an illegal drug in the last month.
2. A survey about the food in the school cafeteria was conducted by passing out questionnaires to students as they entered the cafeteria. A drop box for completed forms as on a table by the cash register .
3. The magazine Harley Davidson Today sent a survey to its subscribers asking whom they admire most in America.
4. A poll of parents in Texas found that 90% of parents say they have spoken to their teenagers about the dangers drinking and driving, while only 45% of those teens say they recall such a discussion.
5. In a census in Russia 1.8 million more women than men reported that they were married.

- SRS
- Stratified
- Cluster
- Systematic
- Multistage
- Convenience

- Undercoverage
- Voluntary Response Bias
- Nonresponse Bias
- Response Bias

6. One year after the Detroit race riots of 1967, interviewers asked a sample of black residents in Detroit if they felt they could trust most white people, some white people, or none at all. When the interviewer was white, 35% answered "most" when the interviewer was black, 7% answered most.

7. A political party mailed questionnaires to all registered voters in Texas, asking whether or not the party should support the death penalty. The voters mailed the completed questionnaires back in an envelope provided.

8. The Nielson rating service estimates the popularity of television stations in the Dallas area. Suppose that four times a year , Nielson takes a random sample of about 5,000 viewers. Every member of the household over age 12 is asked to fill out a diary, showing what he or she watches every quarter hour from 6:00 am to midnight. Each diarist receives $1 for his or her trouble. At the end of one week, Nielson tallies the results from the usable diaries -usually between 33% and 50% of the 5000 sent out.

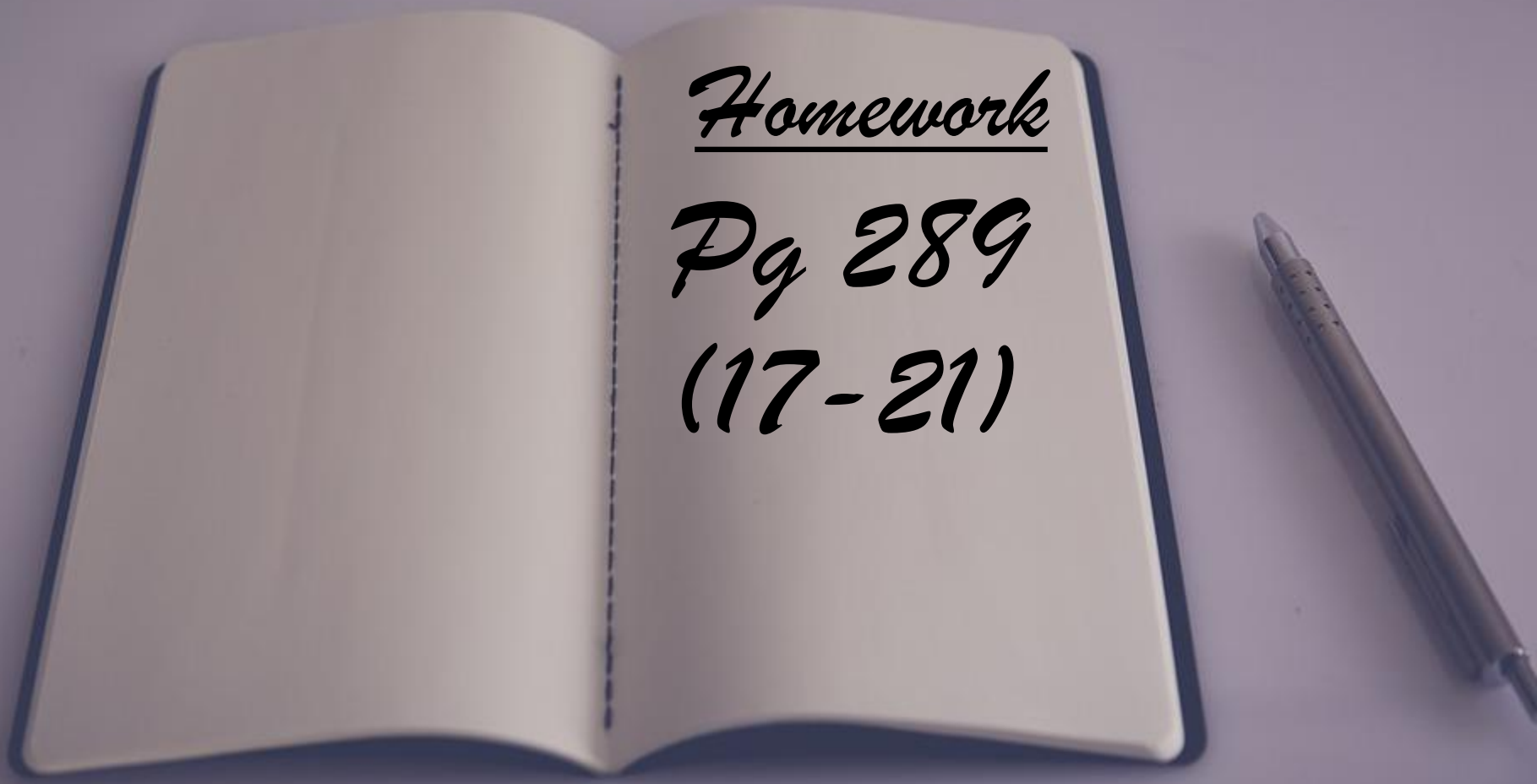
- SRS
- Stratified
- Cluster
- Systematic
- Multistage
- Convenience

- Undercoverage
- Voluntary Response Bias
- Nonresponse Bias
- Response Bias

6. In the 1936 presidential election. Franklin D. Roosevelt ran for reelection against Alfred Landon. As it had done since 1916, The Literary Digest, a popular magazine, ran a pre-election poll. To obtain its sample, the magazine compiled a list of about 10 million names from sources such as telephone books, lists of automobile owners, club membership lists and its own subscription lists. All 10 million people received questionnaires, and about 2.4 million returned them; these people made up the sample. Literary Digest had correctly predicted the winner in all presidential races since 1916. Then in 1936, based on sample responses, the magazine predicted that Landon should win, 57% to 43%. In fact, Roosevelt won, 62% to 38%.

- SRS
- Stratified
- Cluster
- Systematic
- Multistage
- Convenience

- Undercoverage
- Voluntary Response Bias
- Nonresponse Bias
- Response Bias


Homework
Pg 289
(17-21)