

# CHAPTER 4

## Designing Studies

### 4.1b

#### Samples and Surveys

The Practice of Statistics, 5th Edition  
Starnes, Tabor, Yates, Moore



# Samples and Surveys

---

## Learning Objectives

After this section, you should be able to:

- ✓ IDENTIFY the population and sample in a statistical study.
- ✓ IDENTIFY voluntary response samples and convenience samples.  
EXPLAIN how these sampling methods can lead to bias.
- ✓ DESCRIBE how to obtain a random sample using slips of paper, technology, or a table of random digits.
- ✓ **DISTINGUISH a simple random sample from a stratified random sample or cluster sample. Give the advantages and disadvantages of each sampling method.**
- ✓ EXPLAIN how undercoverage, nonresponse, question wording, and other aspects of a sample survey can lead to bias.

# Stratified Random Sample

---

The basic idea of sampling is straightforward: take an SRS from the population and use your sample results to gain information about the population.

Sometimes there are statistical advantages to using more complex sampling methods.

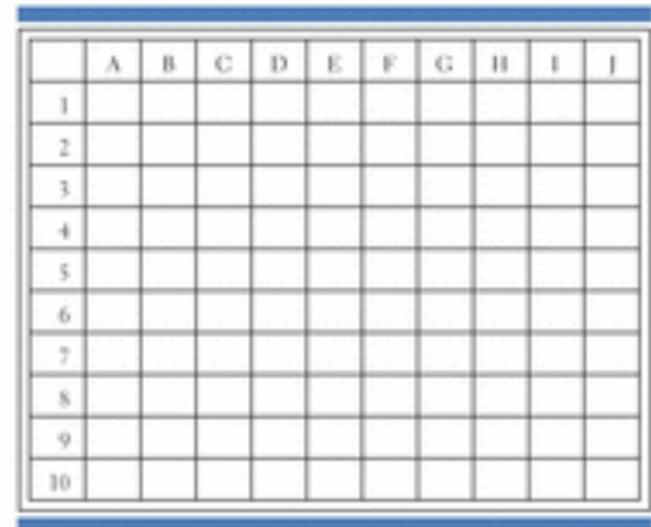
To get a **stratified random sample**, start by classifying the population into groups of similar individuals, called **strata**.

Then choose a separate SRS in each stratum and combine these SRSs to form the sample.

# Sampling Sunflowers

A British farmer grows sunflowers for making sunflower oil. Her field is arranged in a grid pattern, with 10 rows and 10 columns. Irrigation ditches run along the top and bottom of the field. The farmer would like to estimate the number of healthy plants in the field so she can project how much money she'll make from selling them. It would take too much time to count the plants in all 100 squares, so she'll accept an estimate based on a sample of 10 squares.

1. Take a simple random sample of 10 grid squares. Record the location of each square selected.
2. Take a stratified random sample using rows as strata. Record the location of each square selected.
3. Take a stratified random sample using the columns as strata. Record the location of each square.
4. Which method do you believe will give the best overall estimate? Give a justification!

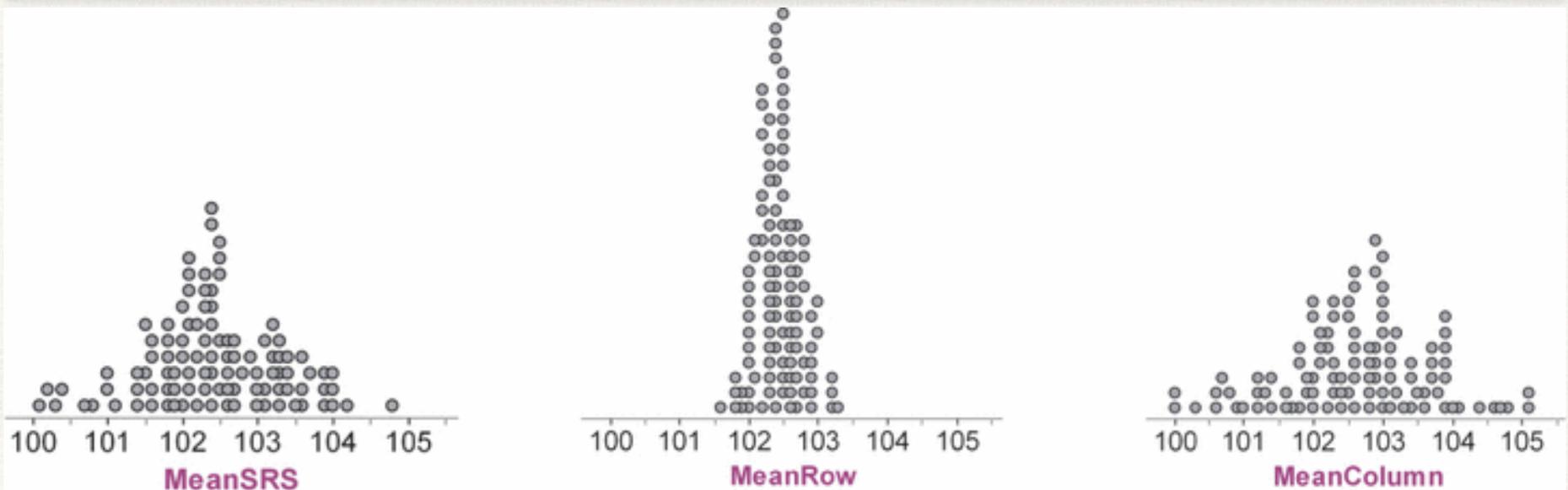


	A	B	C	D	E	F	G	H	I	J
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										

Calculate the mean for each sample and record.

Sampling sunflowers										
	A	B	C	D	E	F	G	H	I	J
1	106	104	106	105	105	105	108	108	107	106
2	103	102	103	104	103	103	103	104	103	105
3	101	99	102	101	100	100	102	101	102	101
4	98	99	99	99	100	99	98	101	99	101
5	98	100	100	100	99	99	99	98	100	98
6	97	98	98	98	99	99	99	99	97	98
7	103	102	102	104	101	102	102	102	104	102
8	104	103	103	103	102	102	103	104	102	102
9	106	106	104	106	102	107	104	103	106	106
10	107	108	109	110	106	107	109	107	106	107

# Sampling Sunflowers



There is much less variability in the estimates using stratified random sampling with rows as strata. Notice in the table that squares within each row had a similar number of healthy plants but there are big differences between rows.

Stratifying by row will give the best estimates because it will take a sample from each different distance from the ditches. In the other two methods, it is possible that squares selected will all be close to the ditches or far from the ditches.

Also note that the true mean of the population is 102.5, and all three distributions of sample means center around 102.5.

# To stratify or not?

---

- When we can choose strata that are “similar within but different between,” stratified random samples give more precise estimates than simple random samples of the same size. (In our Sunflower Activity each row was similar within but different between.)
- Be able to explain how to choose a variable for stratification *and* why stratified random samples are sometimes better than simple random samples.
- Suppose Congress was proposing a massive change to the Social Security program and was going to gather opinions. Which of these variables would be appropriate to use when forming strata?
  - Age
  - Income
  - Gender

# Cluster Sample

---

Although a stratified random sample can sometimes give more precise information about a population than an SRS, both sampling methods are hard to use when populations are large and spread out over a wide area.

In that situation, we'd prefer a method that selects groups of individuals that are "near" one another.

To get a **cluster sample**, start by classifying the population into groups of individuals that are located near each other, called clusters.

Then choose an SRS of the **clusters**. All individuals in the chosen clusters are included in the sample.

# Cluster vs. Strata

---

## Cluster:

- Divide the population into clusters and take all individuals from some clusters (“all from some”)
- Ideally “different within, but similar between”
- Efficiency is the primary benefit (but this is not a convenience sample)
- Ideally each cluster looks like a population

## Strata:

- Divide the population into strata and take some individuals from all strata (“some from all”)
- Ideally “similar within, but different between.”
- Advantage of better information about the population
- Hard to use when populations are large and spread out

You are a “salad quality inspector,” looking for overall quality of the entire salad  
Which type of sampling would be better for:

- a salad bar
- pre-made salads in plastic containers (such as in a fast food restaurant)

# Example – A Good Read

---

A school librarian wants to know the average number of pages in all the books in the library. The library has 20,000 books, arranged by type (fiction, biography, history, and so on) in shelves that hold about 50 books each. You want to select a random sample of 500 books.

1. Explain how to select a simple random sample of 500 books.

Number the books from 1 to 20,000. Use the command `randInt(1,20000)` repeatedly until 500 different numbers between 1 and 20,000 are selected. Use the books that correspond to these numbers.

## Example – A Good Read Cont'd

---

A school librarian wants to know the average number of pages in all the books in the library. The library has 20,000 books, arranged by type (fiction, biography, history, and so on) in shelves that hold about 50 books each. You want to select a random sample of 500 books.

2. Explain how to select a stratified random sample of 500 books. Justify your choice of strata. Why might the librarian want to choose a stratified random sample?

Stratify by type because different types of books might be longer (or shorter) than other types. This will provide a more precise estimate of the average page length than a simple random sample. To select the sample, take an appropriately sized SRS of each type of book and combine the books selected from each type to form the sample. For example, if there are 1000 biographies (5% of 20,000), select an SRS of 25 biographies (5% of 500) using the method described in part (a).

## Example – A Good Read Cont'd

---

A school librarian wants to know the average number of pages in all the books in the library. The library has 20,000 books, arranged by type (fiction, biography, history, and so on) in shelves that hold about 50 books each. You want to select a random sample of 500 books.

3. Explain how to select a cluster sample of 500 books. Justify your choice of clusters. Why might the librarian want to choose a cluster sample?

Clusters are formed by grouping books that are located near each other, making it easier for the librarian. We can use each shelf of 50 books as a cluster and randomly select 10 shelves to obtain the 500 books for our sample. Number the shelves from 1 to 400 and choose an SRS of 10 shelves using the method described in part (a). Then use all the books on the 10 selected shelves as the sample.

# Samples and Surveys

---

## Section Summary

In this section, we learned how to...

- ✓ IDENTIFY the population and sample in a statistical study.
- ✓ IDENTIFY voluntary response samples and convenience samples.  
EXPLAIN how these sampling methods can lead to bias.
- ✓ DESCRIBE how to obtain a random sample using slips of paper, technology, or a table of random digits.
- ✓ **DISTINGUISH a simple random sample from a stratified random sample or cluster sample. Give the advantages and disadvantages of each sampling method.**
- ✓ EXPLAIN how undercoverage, nonresponse, question wording, and other aspects of a sample survey can lead to bias.