2.3 Collecting Samples

This photograph of former U.S. President Harry Truman holding up a copy of the *Chicago Daily Tribune* shortly after the presidential election of 1948 is a classic example of survey results gone wrong. Three different polling agencies had predicted that Thomas E. Dewey would win this election. The fact that Truman has a broad grin on his face tells the real story; he actually won the election by five percentage points. To understand why this happened, you need to learn more about collecting data and creating representative samples.

In this section, you will explore different ways of collecting random samples of data. You will explore sampling techniques and the effect of sample size to learn how to use surveys more effectively.

THE STORY BEHIND THE HEADLINE

Think about
The Headline "Dewey defeats Truman"
How did the pollsters come to their conclusion?

In the United States, the president is indirectly[1] elected by a popular vote. One of two major parties—the Republican Party or the Democratic Party—generally receives in excess of 95% of the votes cast.

The three agencies that conducted polls prior to the 1948 election used a sampling method that required interviewers to choose respondents "carefully." Even though the sample sizes were large (in one case, over 50 000 people) and the sampling technique was designed to elicit responses from a representative cross-section of American voters, the fact that the interviewers chose the respondents was the problem.

TYPES OF SAMPLES

Simple Random Sampling

random—occurring by chance

A simple random sample requires that

- all selections must be *equally likely*
- all combinations of selections must be *equally likely*

A random sample may not end up being representative of the population, but any deviations are due only to chance.

Systematic Random Sampling

sampling interval—found by evaluating $\frac{\text{population size}}{\text{sample size}}$

A systematic random sample is used when you are sampling a fixed percent of the population. A random starting point (i.e., individual, household, or object) is chosen and then you select every *n*th individual for your study, where *n* is the **sampling interval**.

[1] The president is technically elected by the Electoral College. This body of individuals is chosen by each state and the number of representatives is roughly proportional to each state's population. Representatives vote according to the popular vote in their state. This means that most presidential elections reflect the country's overall popular vote.

Stratified Random Sampling

When using a stratified random sample, the population is divided into groups called *strata* (e.g., geographic areas, age groups, places of work, and so on). A simple random sample of the members of each stratum is then taken. The size of the sample for each stratum is proportionate to the stratum's size.

Cluster Random Sampling

Cluster samples require that the population be organized into groups (e.g., schools, communities, companies, and so on). A random sample of groups would then be chosen. All the members of the chosen groups would then be surveyed.

Multi-Stage Random Sampling

Multi-stage samples require that the population be organized into groups. A random sample of groups is chosen and then a random sample of members of the chosen groups is taken.

Destructive Sampling

Samples from which the selected elements cannot be reintroduced into the population are called *destructive sampling* (e.g., light bulbs tested for quality control).

Partial Table of Random Numbers

00	59391	58031
01	99567	76364
02	10363	97518
03	86859	19558
04	11258	24591
05	95068	98628
06	54463	47237
07	16874	62677
08	92494	63157
09	15669	56689
10	99116	75486
11	15696	10703
12	97720	15369
13	11666	13841
14	71628	73130

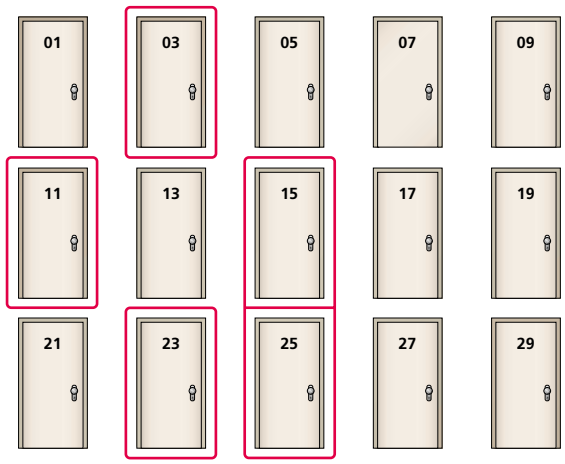
Randomly selected starting digits.

Technolink
For a complete random number table, refer to Appendix B.2 on page 400.

Example 1 Simple Random Sampling

An apartment building superintendent is interested in determining if tenants are satisfied with the maintenance of the building. It is impractical for Mimi to survey every apartment, so she chooses to do a simple random sample.

Solution



For simplicity, this example illustrates how a random sample of 5 units could be selected if there were only 15 apartments in total. In this case, the population consists of the odd-numbered apartments from 1 to 29.

Random numbers can be generated with a calculator or a random number table. A partial table of five-digit random numbers is shown in the margin. To randomly select which columns to use, you could roll a die until you have two distinct values (not including 6). In this table, columns 3 and 4 were randomly selected so that two-digit apartment numbers could be represented. The digits in the square were then randomly

Think about
Random Numbers
Numbers generated by humans cannot be absolutely random. Why?

selected as the starting point. Working down the list, each number that formed an apartment number from the population was noted. In this case, 25, 11, 03, 23, and 15 (ignoring repetition) were the apartments randomly selected.

Remember: A simple random sample requires that

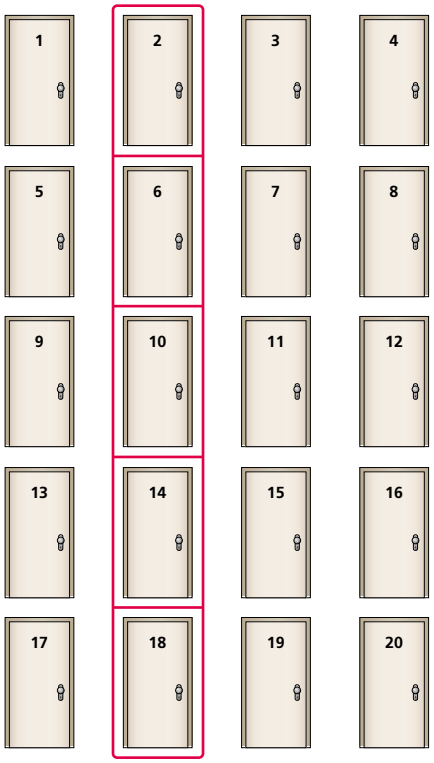
- all selections must be *equally likely*
- all combinations of selections must be *equally likely*

Example 2 Systematic Random Sampling

There are 20 apartments (numbered 1 to 20) on Mimi's floor, and she wants to survey 25% of them. Develop a systematic random sample of the apartments on her floor.

Solution

Mimi must survey 25% of 20 units, or 5 units. To calculate the sampling interval, you divide the population size by the sample size.



$$\frac{\text{population size}}{\text{sample size}} = \frac{20}{5}$$
$$= 4$$

This means that you must sample every fourth apartment. Mimi uses a four-sided die to decide which of the first four apartments she should start at. If Unit 2 is the random starting point, she would then select units 2, 6, 10, 14, and 18.

Systematic random sampling, like simple random sampling, has a random starting point, but it follows a rigid pattern, defined by the sampling interval.

Example 3 Stratified Random Sampling

The Canadian Multicultural Society was interested in urban Canadians' understanding of immigration's impact on the growth of their communities. How can they carry out a stratified random sample?

Solution

They could approach this by sampling and questioning residents in each of four large cities representing different regions of the country. In this case, the strata would be based on four geographic regions of Canada: the Atlantic provinces, Quebec, Ontario, and the Western provinces. Within each stratum, they would select the largest city and then randomly choose adult residents to answer their questionnaire. The number to be sampled in each city would be randomly chosen.

City	Number Sampled
Halifax	200
Montreal	300
Toronto	300
Vancouver	200

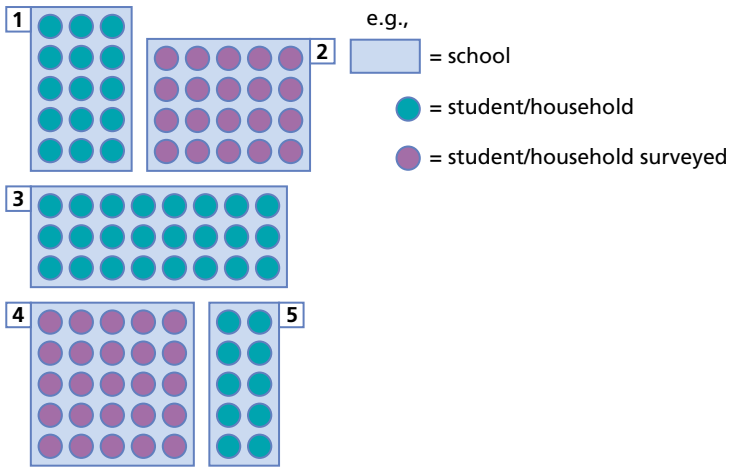
One problem with this method is that Toronto has 10 times the population of Halifax and is underrepresented in this sample. Section 2.5 will further explore avoiding bias in samples.

Example 4 Cluster Random Sampling

The Board of Education for the schools in Lincoln, Ontario, needs to determine parent/guardian opinion about offering summer mathematics courses for elementary school students. Design a cluster random sample the Board can use to select respondents for the survey.

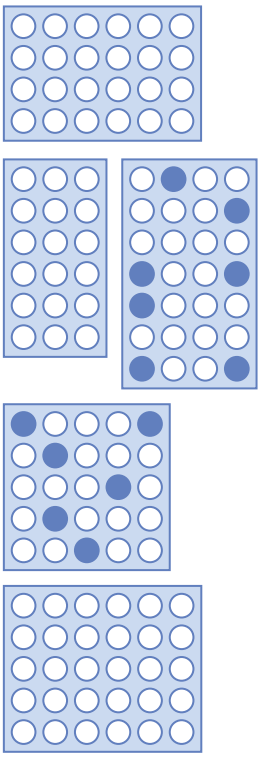
Solution

From the list of five elementary schools in the region, the Board would randomly select two schools. (To do this, they would roll a six-sided die until two of the first five digits appear.). Once the schools are selected, the parents or guardians of each student are surveyed.



Example 5 Multi-Stage Random Sampling

A controversial issue in a school was the banning of all electronic devices from school property. The students' council responded to a few complaints by deciding to collect facts to present to the school administration. Students from the Data Management course suggested that, since this was a very large school, multi-stage sampling would make the most sense. What main steps should they suggest?



Solution

First, they analyzed how much could be accomplished by an agreed deadline given the number of students who could collect data. Their analysis showed that it was feasible to interview 240 students, or 10% of the student body. Because classroom populations vary (age, interest, and so on), they wanted to maximize the number of homerooms represented. However, the administration insisted that a minimum number of homerooms be disrupted. Thus, the students randomly selected 40% of the classes. They then chose 25% of the students to be interviewed (25% of 40% of the classes is 10% of the students).

SAMPLE SIZE: HOW MUCH IS ENOUGH?

It depends! First, the absolute size of the population will have an impact on the appropriate sample size. Furthermore, sample size is related to the reliability of the results. These are some of the factors that will affect reliability:

- the variability of the population (i.e., the more varied the people in the population are, the larger the sample needs to be)
- the degree of precision required for the study
- the sampling method chosen

In practical terms, the larger the sample, the better.

KEY IDEAS

random—occurring by chance; random sampling is used to minimize bias

random sampling techniques—

- *simple random sampling* requires that all selections and combinations of selections are equally likely
- *systematic random sampling* has a random starting point, but follows a pattern defined by the sampling interval; sampling interval is found by dividing the population size by the sample size
- *stratified random sampling* divides the population into groups called *strata*; a simple random sample is taken in each stratum
- *cluster random sampling* organizes the population into groups and then entire groups are randomly selected
- *multi-stage random sampling* organizes the population into groups and then a simple random sample is taken from randomly selected groups
- *destructive sampling* requires that the samples that are taken are destroyed in the process of testing (e.g., cars used in crash tests)

sample size—the larger the sample, the better. Some factors that will affect reliability are the variability of the population, the degree of precision required, and the sampling method chosen.

2.3 Exercises

A

1. **Knowledge and Understanding** Identify the type of random sampling in each of the following scenarios.

(a) The principal randomly selects four classes and surveys each student in those classes.

(b) William picks names out of a hat.

(c) A hockey card collector opens a drawer of sorted cards and, after selecting a random starting point, takes out every fifth card.

(d) The Ministry of Education randomly selects your school for testing, and 40 student names are randomly selected from a student list.

(e) Your class submits solutions to a problem and your teacher divides the work into four piles by achievement levels (Levels 1, 2, 3, and 4). She or he then randomly picks three examples from each.

2. There are 27 students in your class and you wish to interview a random sample of six of them. Describe three practical ways you can select a random sample of this population.

3. State the advantages and disadvantages of each sampling technique listed.

(a) simple random sampling

(b) systematic random sampling

(c) stratified random sampling

(d) cluster random sampling

(e) multi-stage random sampling

4. **(a)** Select 6 pages from this textbook using simple random sampling.

(b) Select 10 pages using systematic random sampling.

(c) Select 12 pages using stratified random sampling.

(d) Select 10 pages using multi-stage random sampling.

5. Based on the following groups of names, identify a sampling method that may have been used to collect the samples listed in parts (a) through (e).

Shaggy	Paul	Joey	Susan
Fred	John	Monica	Elmo
Scooby	George	Rachel	Ernie
Thelma	Ringo	Ross	Oscar
Daphne		Chandler	Zoe
		Phoebe	Maria

(a) Joey, Monica, Fred, Paul, Daphne

(b) Susan, Elmo, Ernie, Oscar, Zoe, Maria

(c) Shaggy, Scooby, Daphne

(d) John, George, Ringo

(e) Shaggy, Fred, George, John, Joey, Chandler, Susan, Ernie

B

6. **(a)** State a technique for selecting a sample of five students from your class using simple random sampling.
(b) Use the table of random numbers or a calculator to select your sample. Repeat this process at least once.
(c) How would you modify your approach if you needed a random sample of eight students consisting of an equal number of males and females?

7. **Communication** Describe the key difference(s) between a systematic random sample and a stratified random sample. For each technique, provide an appropriate example of its use.

8. Describe the steps you would follow to carry out a multi-stage random sample of your school to learn people's opinions about this year's graduation party. What issues would you have to resolve to ensure an accurate result?

9. Design a shuffling method for a standard deck of playing cards that selects a 13-card hand using
(a) simple random sampling
(b) systematic random sampling
(c) stratified random sampling
(d) cluster random sampling
(e) multi-stage random sampling

10. **Application** The following is a list of serial numbers for four types of cell phones.

Model BL	Model BX	Model GB	Model GP
L501	X315	B771	P032
L502	X316	B772	P033
L503	X317	B773	P034
L504	X318	B774	P035
L505	X319	B775	P036
L506	X320	B776	P037
L507	X321	B777	P038
L508	X322	B778	P039

(a) Select 12 phones using simple random sampling.
(b) Select 10 phones using systematic random sampling.
(c) Select 12 phones using stratified random sampling.

11. For Question 10, would it make any sense to select a sample of these cell phones using a multi-stage random sampling technique? Why or why not?

12. **Thinking, Inquiry, Problem Solving** For each situation listed below, identify which of the five sampling methods would be most appropriate and explain why.

(a) You require a sample of 30 students from your grade (which consists of approximately 200 students).

(b) There are 12 mixed (female and male) softball teams in a league and opinions are being gathered about changing the league's name and logo. To save time, only a sample of ballplayers may be interviewed.

(c) In a community of about 18 000 people, school uniforms have been proposed and an ad hoc committee of students wants to find out what the whole community thinks about this proposal. The committee can survey 150 people at most.

(d) There are about 3200 houses and apartment units in town and the community leaders have decided that a questionnaire about the recycling program should be conducted in 160 households.

(e) The city wishes to gather opinions on the plans to paint the community centre from this seven groups that use the building.

C **13.** Do some research to determine what sampling method was used by the three polling agencies (Crossley, Gallup, and Roper) in the story at the beginning of this section. Why did this method produce misleading results?

Technolink
For help on performing simulations with Fathom™, see Appendix D.9 on page 423.

14. Use Fathom™ software to simulate the effect of increasing sample size on the accuracy of results for a given population. From a known population of 1000, take the different sized samples listed below and compare them. As the sample size increases, what do you observe?

first sample	25	second sample	50
third sample	100	fourth sample	250
fifth sample	500	sixth sample	950

ADDITIONAL ACHIEVEMENT CHART QUESTIONS

15. **Knowledge and Understanding** Daily production of brakes at an auto-parts manufacturer is 20 000 units. Suggest a sampling method that would enable the company to check the quality of daily production. How many items should be considered in the sample?

16. **Application** From a list of 100 Grade 12 students numbered 00 to 99, a sample of five is taken. For each example below, identify what sampling method was used. Justify your choices.

(a) 17, 37, 57, 77, and 97 **(b)** 05, 17, 52, 61, and 88

(c) 12, 34, 43, 75, and 90 **(d)** 21, 22, 83, 84, and 85

17. **Thinking, Inquiry, Problem Solving** All of the sampling techniques discussed in this section are examples of probability sampling where *all* members of the population have an equal chance of being selected. What is a non-probability sample?

18. **Communication**

(a) Explain the difference between cluster random sampling and stratified random sampling.

(b) Why do research organizations such as the Gallup organization use stratified random sampling when conducting polls for political elections?

Chapter Problem

Mystery Most Mathematical—Part III

Subject: Re: Puzzle - Part 3
Date: Saturday 04:55:41 -0700 (EDT)
From: C=2*314159*?@homework.com
To: jto@coldmail.com

Suppose that you wished to determine if high school students' opinions about the quality of a movie are related to their gender or age. You only have enough time to carry out 60 interviews. Define the population for the study. Propose a sampling method and explain why you chose it. Make sure that you provide complete details about the procedures that you would use to choose the sample.

jto@coldmail.com aside:
This is the third communication that I have received and I still haven't got a clue who is writing to me. Each time I have received a message, I reply immediately, but all my replies bounce back. The sender is clearly covering her or his tracks. There must be a pattern! It's so frustrating that I can't see it yet.