Section 1.3 – Sampling Methods

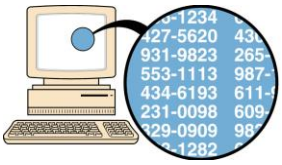
Obtain a Simple Random Sample

Definitions

Random sampling is the process of using chance to select individuals from a population to be included in the sample.

If convenience is used to obtain a sample, the results of the survey are meaningless.

A selection so that each individual member has an equal chance of being selected



A sample of size n from a population of size N is obtained through ***simple random sampling*** if every possible sample of size n has an equally likely chance of occurring. The sample is then called a ***simple random sample.***

Probability Sample selecting members from a population in such a way that each member of the population has a known (but not necessarily the same) chance of being selected

Example

Suppose a study group of consists of 5 students: Bob, Patricia, Mike, Jan, and Maria
2 of the students must go to the board to demonstrate a homework problem. List all possible samples of size 2 (without replacement).

Solution

- Bob, Patricia
- Bob, Mike
- Bob, Jan
- Bob, Maria
- Patricia, Mike
- Patricia, Jan
- Patricia, Maria
- Mike, Jan
- Mike, Maria
- Jan, Maria

Example

Sophia has four tickets to a concert Six of her friends, Yolanda, Michael, Kevin, Marissa, Annie, and Katie, have all expressed an interest in going to the concert. Sophia decides to randomly select three of her six friends to attend the concert.

a) List all possible samples of size $n = 3$ from the population of size $N = 6$. Once an individual is chosen, he or she cannot be chosen again.

b) Comment on the likelihood of the sample containing Michael, Kevin, and Marissa.

Solution

a) The possible samples of size 3 are:

Yolanda, Michael, Kevin	Yolanda, Michael, Marissa	Yolanda, Michael, Annie	Yolanda, Michael, Katie
Yolanda, Kevin, Marissa	Yolanda, Kevin, Annie	Yolanda, Kevin, Katie	Yolanda, Kevin, Annie
Yolanda, Marissa, Katie	Yolanda, Annie, Katie	Michael, Kevin, Marissa	Michael, Kevin, Annie
Michael, Kevin, Katie	Michael, Marissa, Annie	Michael, Marissa, Katie	Michael, Annie, Katie
Kevin, Marissa, Annie	Kevin, Marissa, Katie	Kevin, Annie, Katie	Marissa, Annie, Katie

There are 20 possible sample size 3 from population of 6.

b) Only 1 of the 20 possible samples contains Michael, Kevin, and Marissa, so there is a 1 in 29 chance that the simple random sample will contain these three. In fact, all the samples of size 3 have a 1 in 20 chance of occurring.

Steps for Obtaining a Simple Random Sample

1. Obtain a frame that lists all the individuals in the population of interest. Number the individuals in the frame 1 – *N*.
2. Use a random number table, graphing calculator, or statistical software to randomly generate *n* numbers where *n* is the desired sample size.

Example

The 112th Congress of the United States had 435 members in the House of Representatives. Explain how to conduct a simple random sample of 5 members to attend a Presidential luncheon. Then obtain the sample.

Solution

***Step* 1** Put the members in alphabetical order. Number the members from 1 - 435.

***Step* 2** Randomly select five numbers using a random number generator. First, set the seed. The seed is an initial point for the generator to start creating random numbers—like selecting the initial point in the table of random numbers. The seed can be any nonzero number. Then generate the random numbers. Match the generated random numbers to the corresponding Representatives.

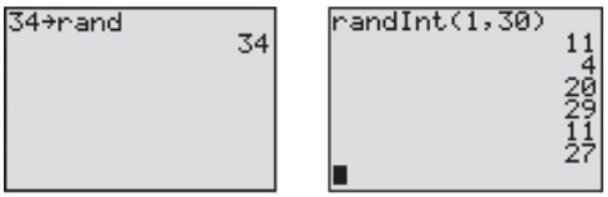
Example

The accounting firm of Sense and Associates has grown. To make sure their clients are still satisfied with the services that are receiving, the company decides to send a survey out to a simple random sample of 5 of its 30 clients. Find a simple random sample of five clients.

Solution

***Step* 1**: We can create a client list and number them from 01 to 10.

***Step* 2**: A table of random numbers can be used to select the individuals to be in the sample.



The following numbers are: 11, 4, 20, 29, 11, 27
We ignore the second 11 because we are sampling without replacement. The clients corresponding to these numbers are the clients to be surveyed.

2. Obtain a Stratified Sample

Definition

A ***stratified sample*** is one obtained by separating the population into homogeneous, nonoverlapping groups called strata, and then obtaining a simple random sample from each stratum. The individuals within each stratum should be homogeneous (or similar) in some way.

Stratified Sampling subdivide the population into at least two different subgroups that share the same characteristics, then draw a sample from each subgroup (or stratum)

Example

In 2008, the United States Senate had 47 Republicans, 51 Democrats, and 2 Independents. The president wants to have a luncheon with 4 Republicans, 4 Democrats and 1 Other. Obtain a stratified sample in order to select members who will attend the luncheon.

To obtain the stratified sample, conduct a simple random sample within each group. That is, obtain a simple random sample of 4 Republicans (from the 47), a simple random sample of 4 Democrats (from the 51), and a simple random sample of 1 Other from the 100. Be sure to use a different seed for each stratum.

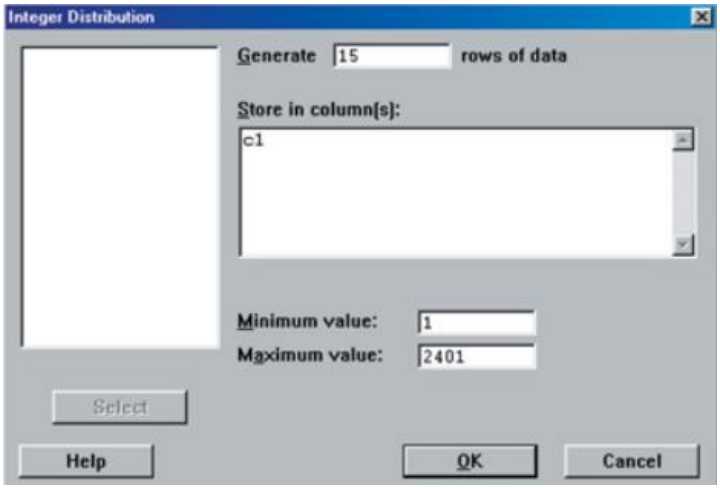
Example

The president of DePaul University wants to conduct a survey to determine the community's opinion regarding campus safety. The president divides the DePaul community into three groups: resident students,, nonresident (commuting) students, and staff (including faculty) so that he can obtain a stratified sample, Suppose there are 6204 resident students, 13,304 nonresident students, and 2401 staff, for a total of 21,909 individuals in the population. The president wants to obtain a sample of size 100, with the number of individuals selected from each stratum weighted by the population size. So resident students

make up $\frac{6204}{21{,}909} = 28\%$, nonresident students account for 61% and staff constitute 11% of the sample. A sample of size 100 requires a stratified sample of $0.28(100) = 28$ resident students, $0.61(100) = 61$ nonresident, and $0.11(100) = 11$ staff.

Solution

Using MINITAB, with the seed set to 4032 and the values, we can obtain the following sample of staff: 240, 230, 847, 190, 2096, 705, 2320, 323, 701, 471, 744



Repeat this proceed for the resident and nonresident students using a different seed.

3. Obtain a Systematic Sample

Definition

A ***systematic sample*** is obtained by selecting every k^{th} individual from the population. The first individual selected is a random number between 1 and *k*.

Systematic Sampling Select some starting point and then select every *k*th element in the population

Steps in Systematic Sampling

Step 1: Determine the population size, *N*.
Step 2: Determine the sample size desired, *n*.
Step 3: Compute *N*/*n* and round down to the nearest integer. This value is *k*.
Step 4: Randomly select a number between 1 and *k*. Call this number *p*.
Step 5: The sample will consist of the following individuals:

$$p,\ p+k,\ p+2k,\ \ldots,\ p+(n-1)k$$

Example

A quality control engineer wants to obtain a systematic sample of 25 bottles coming off a filling machine to verify the machine is working properly. Design a sampling technique that can be used to obtain a sample of 25 bottles.

Solution

If we choose a number between 1 and 7, say 5. Then

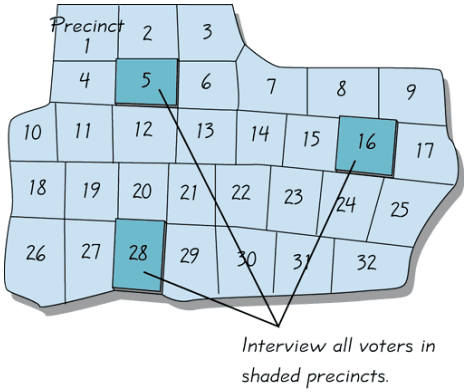
5, $5+7=12$, 19, 26, …, 173 $\qquad\qquad$ $p,\ p+k,\ p+2k,\ \ldots,\ p+(n-1)k$

4. Obtain a Cluster Sample

Definition

A ***cluster sample*** is obtained by selecting all individuals within a randomly selected collection or group of individuals.

Cluster Sampling divide the population area into sections (or clusters); randomly select some of those clusters; choose all members from selected clusters



Multistage Sampling Collect data by using some combination of the basic sampling methods

In a multistage sample design, pollsters select a sample in different stages, and each stage might use different methods of sampling.

Example

A sociologist wants to gather data regarding household income within the city of Boston. Obtain a sample using cluster sampling

Solution

Suppose there are 10,493 city blocks in Boston. First, the sociologist must number the blocks from 1 to 10,493. Suppose the sociologist has enough time and money to survey 20 clusters (city blocks). The sociologist should obtain a simple random sample of 20 numbers between 1 and 10,493 and survey all households from the clusters selected. Cluster sampling is a good choice in this example because it reduces the travel time to households that is likely to occur with both simple random sampling and

stratified sampling. In addition, there is no need to obtain a frame of all the households with cluster sampling. The only frame needed is one that provides information regarding city blocks

Example

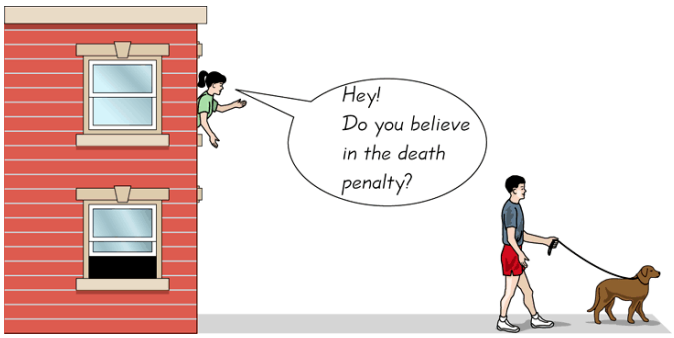
The U.S. government's unemployment statistics are based on surveyed households. It is impractical to personally visit each member of a simple random sample, because individual households would be spread all over the country, Instead, the U.S. Census Bureau and the Bureau of Labor Statistics combine to conduct a survey called the Current Population Survey. This survey obtains data describing such factors as unemployment rates, college enrollments, and weekly earning amounts. The survey incorporates a multistage sample design, roughly following these steps:

1. The surveys partition the entire United States into 2007 different regions called *primary sampling units* (PSU). The primary sampling units are metropolitan areas, large counties, ot groups of smaller counties.
2. The surveyors select a sample of primary sampling units in each of the 50 states. For the Current Population Survey, 792 of the primary sampling units are used. (All of the 432 primary sampling units with the largest populations are used, and 360 primary sampling units are randomly selected from the other 1575.)
3. The surveyors partition each of the 792 selected primary sampling units into blocks, and they then use stratified sampling to select a sample of blocks.
4. In each selected block, surveyors identify clusters of households that are close to each other. They randomly select clusters, and they interview all households in the selected clusters.

This multistage sample design includes random, stratified, and cluster sampling at different stages. The end result is a complicated sampling design, but it is much more practical and less expensive than using a simpler design, such as using a simple random sample.

Definition

A ***convenience sample*** is one in which the individuals in the sample are easily obtained.
Convenience Sampling use results that are easy to get



Any studies that use this type of sampling generally have results that are suspect. Results should be looked upon with extreme skepticism.

Example

In practice, most large-scale surveys obtain samples using a combination of the techniques just presented. As an example of multistage sampling, consider Nielsen Media Research. Nielsen randomly selects households and monitors the television programs these households are watching through a People Meter. The meter is an electronic box placed on each TV within the household. The People Meter measures what program is being watched and who is watching it. Nielsen selects the households with the use of a two-stage sampling process.

Solution

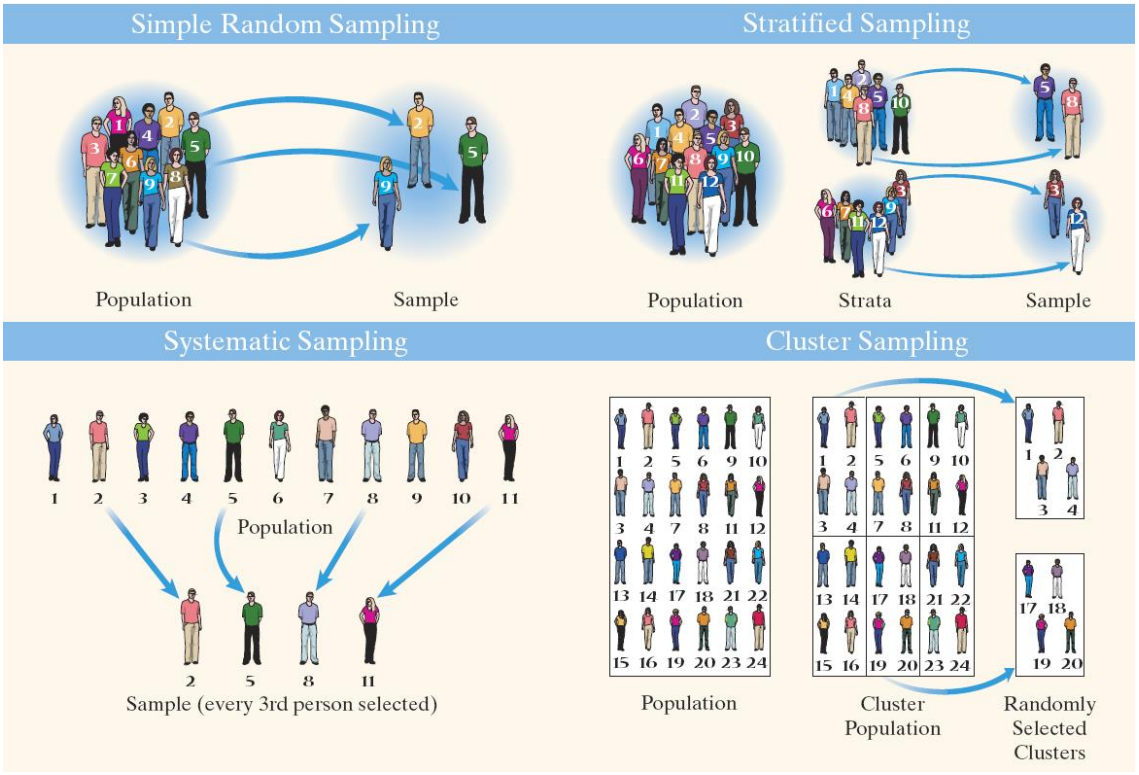
***Stage* 1** Using U.S. Census data, Nielsen divides the country into geographic areas (strata). The strata are typically city blocks in urban areas and geographic regions in rural areas. About 6000 strata are randomly selected.

***Stage* 2** Nielsen sends representatives to the selected strata and lists the households within the strata. The households are then randomly selected through a simple random sample.

Nielsen sells the information obtained to television stations and companies. These results are used to help determine prices for commercials.

As another example of multistage sampling, consider the sample used by the Census Bureau for the Current Population Survey. This survey requires five sample stages of sampling:

***Stage* 1**: Stratified sample
***Stage* 2**: Cluster sample
***Stage* 3**: Stratified sample
***Stage* 4**: Cluster sample
***Stage* 5**: Systematic sample



Exercises *Section* 1.3 – Sampling Methods

1. A student of the author collected measurements of arm lengths from her family members. Identify what type is used: random, systematic, convenience, stratified, or cluster.

2. On the day of the last presidential election, ABC News organized an exit poll in which specific polling stations were randomly selected and all voters were surveyed as they left the premises. Identify what type is used: random, systematic, convenience, stratified, or cluster.

3. The author was an observer at a town of Poughkeepsic Police sobriety checkpoint at which every fifth driver was stopped and interviewed. (He witnessed the arrest of a former student.) Identify what type is used: random, systematic, convenience, stratified, or cluster.

4. You observed professional wine taster working at the Consumer's Union testing facility in NY. Assume that a taste test involves three different wines randomly selected from each of five different wineries. Identify what type is used: random, systematic, convenience, stratified, or cluster.

5. The U.S. Department of Corrections collects data about returning prisoners by randomly selecting five federal prisons and surveying all of the prisoners in each of the prisons. Identify what type is used: random, systematic, convenience, stratified, or cluster.

6. You instructor surveyed all of his students to obtain sample consisting of the number of credit cards students possess. Identify what type is used: random, systematic, convenience, stratified, or cluster.

7. In a study of college programs, 820 students are randomly selected from those majoring in communications, 1463 students are randomly selected from those majoring in business, and 760 students are randomly selected from those majoring in history. Identify what type is used: random, systematic, convenience, stratified, or cluster.

8. Pharmacists typically fill prescriptions by scooping a sample of pills from a larger batch that is in stock. A pharmacist thoroughly mixes a large batch of Lipitor pills, then selects 30 of them. Does this sampling plan result in a random sample? Simple random sample? Explain.

9. A quality control engineer selects every $10{,}000^{th}$ M&M plain candy that is produced. Does this sampling plan result in a random sample? Simple random sample? Explain.

10. NBC News polled reactions to the last presidential election by surveying adults who were approached by a reporter at a location in N.Y. City. Does this sampling plan result in a random sample? Simple random sample? Explain.

11. A classroom consists of 36 students seated in six different rows, with six students in each row. The instructor rolls a die to determine a row, then rolls the die again to select a particular student in the row. This process is repeated until a sample of 6 students is obtained. Does this sampling plan result in a random sample? Simple random sample? Explain.

12. A computer company employs 100 software engineers and 100 hardware engineers. The personnel manager randomly selects 20 of the software engineers and 20 of the hardware engineers and questions them about career opportunities within the company. Does the sampling plan result in a random sample? Simple random sample? Explain.

13. A polling company obtains an alphabetical list of names of voters in a precinct. They select every 20^{th} person from the list until a sample of 100 is obtained. They then call these 100 people. Does the sampling plan result in a random sample? Simple random sample? Explain.

14. What is an inherent zero? Describe three examples of data sets that have inherent zeros and three that do not.

15. What is the different between a random sample and a simple random sample?

16. Determine whether the statement is true or false. If false, rewrite it as a true statement

a) In a randomized block design, subjects with similar characteristics are divided into blocks, and then, within each block, randomly assigned to treatment groups.

b) Using a systematic sample guarantees that members of each group within a population will be sampled.

c) The method for selected a stratified sample is to order a population in some way and then select members of the population at regular intervals.

17. Which method of data collection should be used to collect data for the following study

a) A study of the health of 148 kidney transplant patients at a hospital.

b) A study of the effect on the taste of a snack food made with a sugar substitute

c) A study of how fast a virus would spread in a herd of cattle.

18. A pharmaceutical company wants to test the effectiveness of a new allergy drug. The company identifies 250 females 30-35 years old who suffer from severe allergies. The subjects are randomly assigned into two groups. One group is given the new allergy drug and the other is given a placebo that looks exactly like the new allergy drug. After six months, the subjects' symptoms are studied and compared

a) Identify the experimental units and treatment used in this experiment.

b) Identify a potential problem with the experiment design being used and suggest a way to improve it.

c) How could this experiment be designed to be a double-blind?

19. What type of sampling is used: random, stratified, convenience, cluster, systematic, in the following?

a) To estimate the percentage of defects in a recent manufacturing batch, a quality-control manager at Intel selects every 8^{th} chip that comes off the assembly line starting with the 3^{rd} until she obtains a sample of 140 chips.

b) To determine the prevalence of human growth hormone (HGH) use among high school varsity baseball players, the State Athletic Commission randomly selects 50 high schools. All members of the selected high schools' varsity baseball teams are tested for HGH.

c) To determine customer opinion of its boarding policy. Southwest Airlines randomly selects 60 lights during a certain week and surveys all passengers on the flights.

d) A member of Congress wishes to determine her constituency's opinion regarding estate taxes. She divides her constituency into three income classes: low-income households, middle-income households, and upper-income households. She then takes a simple random sample of households from each income class.

e) In an effort to identify whether an advertising campaign has been effective, a marketing firm conducts a nationwide poll by randomly selecting individuals from a list of known users of the product.

f) A radio station asks it listeners to call in their opinion regarding the use of U.S forces in peacekeeping missions.

g) A farmer divides his orchard into 50 subsections, randomly selects 4, and samples all the trees within the 4 subsections to approximate the yield of this orchard.

h) A college official divides the student population into five classes: freshman, sophomore, junior, and graduate student. The official takes a simple random sample from each class and asks the members' opinions regarding student services.

i) Toyota wants to administer a satisfaction survey to its current customers. Using their customer database, the company randomly selects 80 customers and asks them about their level of satisfaction with the company

j) To determine her power usage, Dan divides up his day into three parts: morning, afternoon, and evening. He then measures his power usage at 3 randomly selected times during each part of the day.

k) A newspaper asks its readers to call in their opinion regarding the number of books they have read this month.

l) Toshiba wants to administer a satisfaction survey to its current customers. Using their customer database, the company randomly selects 80 customers and asks them about their level of satisfaction with the company.

m) An education researcher randomly selects 48 middle schools and interviews all the teachers at each school.

n) A market researcher selects 500 drivers under 30 years of age and 500 drivers over 30 years of age.

o) To avoid working late, a quality control analyst simply inspects the first 100 items produced in a day.

20. Determine whether you would take a census or use a sampling to collect data for the study described:

a) The average credit card debt of the 65 employees of a company

b) The most popular grocery store among the 40,000 employees of a company

21. Determine if the survey question is biased. If the question is biased, suggest a better wording

a) Why drinking fruit juice good for you?

b) Why is eating ice cream bad for you?

22. A company has been rating television programs for more than 60 years. It uses several sampling procedures, but its main one is to track the viewing patterns of 20,000 households. These contain more than 45,000 people and are chosen to form a cross section of the overall population. The

households represent various locations, ethnic groups, and income brackets. The data gathered from the sample of 20,000 households are used to draw inferences about the population of all households in the U.S.

a) What strata are used in the sample?
b) Why is it important to have a stratified sample for these ratings?
c) Observation studies are sometimes referred to as natural experiments. Explain what this means

23. Some polling agencies ask people to call a telephone number and give their response to a question
a) What is an advantage of this type of survey?
b) What is disadvantage of this type of survey?
c) Identify the sampling technique used.

24. A computer company employs 100 software engineers and hardware engineers. The personnel manager randomly selects 20 of the software engineers and 20 of the hardware and questions them about career opportunities within the company. Does this sampling plan result in a random sample? Simple random sample? Explain.

25. Suppose you are the president of the student government. You wish to conduct a survey to determine that student body's opinion regarding student services. The administration provides you with a list of the names and phone numbers of the 19,935 registered students.
a) Discuss the procedure you would follow to obtain a simple random sample of 25 students.
b) Obtain this sample

26. True or False
a) When taking a systematic random sample of size n, every group of size n from the population has the same chance of being selected
b) A simple random sample is always preferred because it obtains the same information as other sampling plans but requires a smaller sample size.
c) When conducting a cluster sample, it is better to have fewer cluster with more individuals when the clusters are heterogeneous.
d) Inferences based on voluntary response samples are generally not reliable.
e) When obtaining a stratified sample, the number of individuals included within each stratum must be equal.

27. The human resource department at a certain company wants to conduct a survey regarding worker morale. The department has an alphabetical list of all 4502 employees at the company and wants to conduct a systematic sample.
a) Determine k if the sample size is 50
b) Determine the individuals who will be administered the survey. More than one answer is possible.

28. To predict the outcome of a county election, a newspaper obtains a list of all 945,035 registered voters in the county and wants to conduct a systematic sample.
a) Determine k if the sample size is 130
b) Determine the individuals who will be administered the survey. More than one answer is possible.