Interactive
Statistics
Third Edition

Martha Aliaga

Brenda Gunderson

Chapter 2
Producing Data

1. Fill in the blank. A _____ is a numerical value calculated from all the units in the population.
   (a) parameter
   (b) statistic
   (c) variable

2. Fill in the blank. A _____ is a numerical value calculated from all the units in the sample.
   (a) parameter
   (b) statistic
   (c) variable

3. Fill in the blank. A _____ is a sample consisting of the entire population.
   (a) population size
   (b) survey
   (c) census

4. Fill in the blank. The size of the population is denoted by _____.
   (a) n
   (b) N
   (c) p

5. A convenience sample is a form of a simple random sample.
   (a) True
   (b) False

Questions 3 and 4 are based on the following scenario.
   In Wisconsin 85% of cats and dogs are neutered. In a sample of 350 cats and dogs from throughout the state, only 212 were neutered for a percentage of 60.6%.

6. Fill in the blank. In this situation, the value 60.6% is a _____ .
   (a) parameter
   (b) statistic
   (c) variable

7. Fill in the blank. In this situation, the neuter status of the cats and dogs is a _____.
   (a) parameter
   (b) statistic
   (c) variable

8.   Fill in the blank. _____ is the systematic tendency on the part of the sampling procedure to include or exclude a certain type of unit.
        (a)   Response bias
        (b)   Nonresponse bias
        (c)   Selection bias

9.   Fill in the blank. _____ is the sampling method that first divides the population into groups and then a simple random sample is taken from each group.
        (a)   Cluster sampling
        (b)   Multistage sampling
        (c)   Stratified sampling
        (d)   Systematic sampling

10.   Fill in the blank. _____ is the sampling method that first divides the population into groups and then a simple random sample of groups is taken.
        (a)   Cluster sampling
        (b)   Multistage sampling
        (c)   Stratified sampling
        (d)   Systematic sampling

11.   The actual proportion of all chemistry majors at your university is a statistic.
        (a)   True
        (b)   False

12.   In stratified random sampling, if all strata contain the same number of units, then stratified random sampling is the same as simple random sampling.
        (a)   True
        (b)   False

13.   Mutually exclusive subgroups imply that each unit of the population belongs to one and only one subgroup.
        (a)   True
        (b)   False

14.   For a 1-in-k systematic sample, we order the units of the population in some way and randomly select one unit from the population as the first unit in the sample. We continue by taking the next k units to complete the sample.
        (a)   True
        (b)   False

15.   In a 1-in-3 systematic sampling, one cannot obtain a sample with two consecutive units of the population.
        (a)   True
        (b)   False

16.   For cluster sampling to be effective, the units of the population should be divided up into homogeneous groups.
       (a)   True
       (b)   False

17.   A cluster sample is a simple random sample.
       (a)   True
       (b)   False

18.   In cluster sampling, the chance that any unit is selected depends on the size of the cluster it belongs to.
       (a)   True
       (b)   False

19.   If a population is divided into groups for which the units within a group are relatively homogeneous, then it is preferable to take a stratified sample rather than a cluster sample from the population.
       (a)   True
       (b)   False

20.   In a probability sample each unit in the population has the same chance as any other unit in the population of being in the sample.
       (a)   True
       (b)   False

21.   A researcher would like to estimate the proportion of adult voters who are in favor of Proposition A. The population of adult voters is stratified into males and females. Sixty percent of the population is known to be male. A stratified random sample of size 100 (50 males and 50 females) is taken from the population. If the sample proportion of males favoring Proposition A is 0.20, while the sample proportion of females favoring Proposition A is 0.70, then estimate the proportion of adult voters in the population favoring Proposition A.

22.   A professor was interested in obtaining information about the different part-time jobs university students hold. His sample consists of all students who donated blood at the University blood drive campaign. Name this sampling method.

23.   Some television stations take quick polls of public opinion by announcing a question on the air and asking viewers to call one of two telephone numbers to register their opinion as "Yes" or "No". Telephone companies make available "900" numbers for this purpose; dialing such a number results in a small charge to your telephone bill. One such call-in poll finds that 73% of those who called are opposed to a proposed local gun control ordinance. Give one reason why this sampling method can be biased.

24.   An education researcher wants to study a new early childhood instruction method and approaches the three preschools located near her university to recruit their students for her study. Explain why this sampling method can be biased.

25. Jose and Marie are taking a statistics class. They wish to estimate the proportion of students in their class who own a car. If they wish to estimate this proportion separately for each class level (freshmen, sophomores, juniors, and seniors), what sampling method should they use?

26. Farmer Bob has a small apple orchard on his farm. The trees are nicely lined up in 8 rows of 10 trees. He randomly selects 2 of these rows and sells all the apples from the trees in these two rows. What sampling method did he use?

Questions 27 through 30 are based on the following scenario.
A claim is made that 60% of the all undergraduate students at a community college have their own computer. Amy thinks this figure is too high and decides to collect some data to test her hypothesis. She has a list of all 2400 students enrolled at the college this term and wishes to sample 5% of these students.

27. How many students does she plan to have in the sample?

28. Using your TI calculator with a seed value of 28, give the labels for the first 10 students to be selected for the sample.

29. Suppose the students in the list are only the full-time students enrolled at this community college. Part-time and other special category students would not have a chance to be included in the sample. What type of bias is this an example of?

30. Amy finds that 50% of the students at this college have their own computer. The school newspaper writes about her findings, saying "50% of college students own a computer." What is wrong with this conclusion.

Questions 31 through 36 are based on the following scenario.
Teacher Emily wishes to estimate the number of female students in her class of size 40 using cluster sampling. The students are seated in 10 rows numbered 1, 2, 3, ..., 10. The first 4 rows are shown below.

Row 1	Ann	Jill	John	Amy	
Row 2	Mark	Frank	Annabella		
Row 3	Victoria	Sally	Joe	Martha	Charlie
Row 4	David	Jack		Isabella	

There are <u>no</u> students in the class with the same first names. With rows as clusters, simple random sampling will be used to select 1 cluster from the total of 10 clusters.

31. What is the chance that Sally and Joe will be included in the cluster sample?

32. What is the chance that David will not be included in the cluster sample?

33. What is the chance that Jill and Frank will be included in the cluster sample?

34. Using your TI calculator with a seed value of 235, select 1 cluster at random from the population of clusters. Give the selected row number, the number of students in the selected row, and the number of female students in the selected row.

35. Can you determine the sample proportion of female students, $\hat{p}$? If yes, compute it. If no, explain why not.

36. Can you determine the population proportion of female students, p? If yes, compute it. If no, explain why not.

Questions 37 through 40 are based on the following scenario.
A sociologist wanted to know which academic subject high school students preferred. The hypotheses he wishes to test are: H_0: The percentage of female high school students who prefer mathematics is equal to the percentage of male students who prefer mathematics, versus H_1: The percentage of female high school students who prefer mathematics is lower than the percentage of male students who prefer mathematics. Of the 2000 high school students, 1105 were males and 895 were females. The sociologist randomly selected 100 of the female high school students and 100 of the male high school students, and asked the 200 students what academic subject they prefer. Among the 100 female students, 20% preferred mathematics; among the 100 male students, 30% preferred mathematics.

37. What sampling method was used to select the 200 students?

38. Using your TI calculator with a seed of 32, identify the labels for the first 5 female students selected.

39. Suppose the p-value for this test is 0.03. Is the difference in percentages significant at a 5% level? Explain.

40. Suppose the p-value for this test is 0.03. Is the difference in percentages significant at 1% level? Explain.

Questions 41 through 44 are based on the following scenario.
On the north side of Birch Lane there are 43 houses. Jeff, who lives on the south side of Birch Lane, wishes to estimate the proportion of families on the north side that own one or more pets. He will use a 1-in-5 systematic sampling method to obtain his sample. The house farthest east will be house 1 and the house farthest west will be house 43.

41. If the first house in the sample is house 2, how many houses will be included in the sample?

42. If the first house in the sample is house 4, how many houses will be included in the sample?

43. Using your TI calculator with a seed value of 35, give the labels of all houses in the sample.

44.   Suppose that at exactly 3 of the houses sampled in question 37 the families own one or more pets. What is the sample proportion of pet owners, $\hat{p}$?

Questions 45 through 49 are based on the following scenario.
Oak Park is located in three different counties. The wooded area appears to be infected with a disease. The park ranger wishes to estimate the proportion of infected trees. Of the 1065 trees in the park, 405 are located in Abel County, 300 in Bedford County, and the remaining 360 in Carlisle County. The park ranger would like to ensure that trees from each county will be sampled.

45.   What sampling method should the park ranger use?

46.   The park ranger decides to sample 20% of the trees from each county. How many trees will be sampled from Carlisle County?

47.   How many trees will be sampled overall?

48.   Of the trees sampled from Abel County, 12 turn out to be infected with the disease. What is the sample proportion, $\hat{p}$, of infected trees in Abel County?

49.   Of the trees sampled in Bedford County 10% are infected with the disease and of the trees sampled in Carlisle County 12.5% are infected. What is the overall weighted estimate of the proportion of infected trees?

Questions 50 through 57 are based on the following scenario.
A sociologist in California is interested to find out a little more about the difference in the proportions of people who have quit smoking between people from rural areas and people from the city. She divides the state into 4 geographical zones: North, Mid-North, Mid-South, and South. From each geographical zone, she randomly selects one third of the cities and one third of the rural areas. In each of these selected cities and rural areas she groups people by gender and randomly selects 25 women and 25 men who have ever smoked (or still smoke).

50.   What is the overall sampling method used by the sociologist?

51.   The sociologist used two stages in her procedure. Identify these stages and the units sampled in each stage.

52.   What is the name of the sampling method used in stage 2?

53.   If there are 18 rural areas in each geographical zone, how many groups of 25 rural women will the sociologist end up with?

54.   If there are 15 cities in each geographical zone, what is the total number of people (both from rural areas and from the city) in the sample?

55.   The sociologist would like to differentiate between male and female and between rural and city. How many sample proportions will she calculate?

56.   Of all rural women in the sample, 168 say they have quit successfully (longer than a year since their last cigarette). What is the sample proportion of rural women who have quit successfully?

57.   Of all men in the sample, 290 say they've quit successfully. What is the sample proportion of men who have quit successfully?