
Contents

Chapter 1: Getting Started	3
Chapter 2: Organizing Data	8
Chapter 3: Averages and Variation	31
Chapter 4: Correlation and Regression	53
Chapter 5: Elementary Probability Theory	75
Chapter 6: The Binomial Probability Distribution and Related Topics	97
Chapter 7: Normal Curves and Sampling Distributions	122
Chapter 8: Estimation	152
Chapter 9: Hypothesis Testing.....	175
Chapter 10: Inferences about Differences.....	208
Chapter 11: Additional Topics Using Inference	245

Chapter 1: Getting Started

Section 1.1

1. Individuals are people or objects included in the study, while a variable is a characteristic of the individual that is measured or observed.
2. Nominal data are always qualitative.
3. A parameter is a numerical measure that describes a population. A statistic is a numerical value that describes a sample.
4. If the population does not change, the values of the parameters will not change. Thus, for a fixed population, parameter values are constant. If we take three samples of the same size from a population, the values of the sample statistics will almost surely differ.
5. (a) These numerical assignments are at the nominal level. There is no apparent ordering in the responses.
(b) These numerical assignments are at the ordinal level. There is an increasing relationship from worst to best levels of service. These assignments are not at the interval or ratio level. The distances between numerical responses are not meaningful. The ratios are also not meaningful.
6. Lucy's observations do not apply to all adults; they apply only to her friends. Since the sample is not random, we cannot draw any conclusions about a larger group using this data.
7. (a) Meal ordered at fast-food restaurants.
(b) Qualitative
(c) Response for all U.S. adult fast-food consumers.
8. (a) Miles per gallon.
(b) Quantitative.
(c) All new cars (U.S.)
9. (a) Nitrogen concentration (milligrams of nitrogen per liter of water).
(b) Quantitative.
(c) Nitrogen concentration in the entire lake.
10. (a) Number of ferromagnetic artifacts per 100 square meters.
(b) Quantitative.
(c) The entire Tara region.
11. (a) Ratio. (b) Interval. (c) Nominal. (d) Ordinal. (e) Ratio. (f) Ratio.
12. (a) Ordinal. (b) Ratio. (c) Nominal. (d) Interval. (e) Ratio. (f) Nominal.
13. (a) Nominal. (b) Ratio. (c) Interval. (d) Ordinal. (e) Ratio. (f) Interval.
14. *Form B* is better. Statistical methods can be applied to the ordinal data obtained from *Form B* but not to the open-response answers obtained from *Form A*.
15. (a) Answers vary. Ideally, weigh the packs in pounds using a digital scale that has tenths of pounds for accuracy.
(b) Some students may refuse to have their backpacks weighed.
(c) Informing students before class may cause students to remove items before class.

Section 1.2

In stratified samples, we select a random sample from each stratum. In cluster sampling, we randomly select clusters to be included, and then each member of the cluster is sampled.

1. In simple random samples, every sample of size n has an equal chance of being selected. In a systematic sample, the only possible samples are those including every k th member of the population with respect to the random starting position.
- 2.

Sampling error is the difference between the value of the population parameter and the value of the sample statistic that stems from the random selection process. The term is being used incorrectly here. Certainly larger boxes of cereal will cost more than smaller boxes of cereal.

- 3.
4. The sample frame consists of students who use the college recreation center. No, some students may not use the recreation center.
5. No, even though the sample is random, some students younger than 18 or older than 20 may not have been included in the sample.
6. No, a random sample could include only music majors.
 - (a) Stratified.
 - (b) No, because each pooled sample would have 100 season ticket holders from men's basketball games and 100 for women's basketball games. Samples, for example, with 125 ticket holders for men's basketball games and 75 for women's games are not possible.
8.
 - (a) Yes. Every student has a 50% chance of being selected.
 - (b) It is not possible based on this method of selecting students. Since every sample of size 20 is not possible, this is not a simple random sample.
 - (c) Assign numbers 1, 2, ..., 40 to the students and use a random-digits table or a computer package to draw random numbers.
9. Simply use a random digits table or a computer package to randomly select four students from the class.
 - (a) Answers vary. Perhaps they are excellent students who make an effort to get to class early.
 - (b) Answers vary. Perhaps they are busy students who are never on time to class.
 - (c) Answers vary. Perhaps students in the back row are introverted.
 - (d) Answers vary. Perhaps taller students are healthier.
10.
 - (a) Students who are absent from class on Monday cannot be included in the sample.
 - (b) Home-schooled students, drop out students, or homeless students cannot be included in the sample.
11. Answers vary.
12. Answers vary.
13. Answers vary.
14. Answers vary. One possibility is to use 0, 1, 2, 3, and 4 to indicate heads, and 5, 6, 7, 8, and 9 to indicate tails.
15.
 - (a) Yes, it is appropriate, as a number can repeat itself once it has occurred. The outcome on the fourth roll is 2.
 - (b) We will most certainly not get the same sequence of outcomes. The process is random.
16. Answers vary. We do expect at least one match on birthdays on over 50% of the times we run this experiment.
17. Answers vary. Use single digits on the table to determine the placement of correct answers.
18. Answers vary. The test key would be a random arrangement of *True* and *False* responses.

19. (a) Simple random sampling. Every sample of size n from the population has an equal chance of being selected, and every member of the population has an equal chance of being included in the sample.
 - (b) Cluster sampling. The state, Hawaii, is divided into ZIP Codes. Then, within each of the 10 selected ZIP Codes, all businesses are surveyed.
 - (c) Convenience sampling. This technique uses results or data that are conveniently and readily obtained.
 - (d) Systematic sampling. Every fiftieth business is included in the sample.
 - (e) Stratified sampling. The population was divided into strata based on business type. Then a simple random sample was drawn from each stratum.
20. (a) Stratified sampling. The population was divided into strata (four categories of length of hospital stay), and then a simple random sample was drawn from each stratum.
 - (b) Simple random sampling.
 - (c) Cluster sampling. There are five geographic regions, and some facilities from each region are selected randomly. Then, for each selected facility, all patients on the discharge list are surveyed to create the patient satisfaction profiles.
 - (d) Systematic sampling. Every 500th patient is included in the sample.
 - (e) Convenience sampling.

Section 1.3

1. Answers vary. People with higher incomes will likely have high-speed Internet access, which will lead to spending more time online. Spending more time online might lead to spending less time watching TV. Thus, spending less time watching TV cannot be attributed solely to high income or high-speed internet access.
2. A double-blind procedure would entail neither the patients nor those administering the treatments knowing which patients received which treatments. This process should eliminate potential bias from the treatment administrators and from patient psychology regarding benefits of the drug.
3. No, respondents do not constitute a random sample from the community for several reasons, for instance, the sample frame includes only those at a farmer's market, Jill might not have approached people with large dogs or those who were busy, and participation was voluntary. Jill's T-shirt may have influenced respondents.
4. No, the pooled sample had a fixed number of students from each block.
5. (a) No, those aged 18 – 29 in 2006 became aged 20 – 31 in 2008. The study is looking at the same generation.
- (b) 1977 to 1988, inclusive.
6. By 2016, the Echo generation will be aged 28 – 39, and their perception of items as necessities or luxuries might have changed by then.
7. (a) This is an observational study. The data collection method did not influence the outcome.
- (b) This is an experiment. A treatment was imposed on the sheep in order to prevent heartworm.
- (c) This is an experiment. The restrictions on fishing possibly led to a change in the length of trout in the river.
- (d) This is an observational study. The data was collected without influencing the turtles.
8. (a) Sampling. (b) Simulation. (c) Census. (d) Experiment.
9. (a) Use randomization to select ten calves to inoculate with the vaccine. After a period of time, test all calves for the infection. No placebo is being used.
- (b) Use randomization to select nine schools to visit. After ten weeks, survey students in all 18 schools for their views on police officers. No placebo is being used.
- (c) Use randomization to select 40 subjects to use the skin patch. A placebo is used for the other 35 subjects. At the end of the trial, survey all 75 subjects about their smoking habits.

10. (a) “Over the last few years” could mean 2 years, 3 years, 7 years, etc. A more precise phrase is, “Over the past 5 years.”
- (b) If a respondent is first asked, “Have you ever run a stop sign,” chances are that their response to the question, “Should fines be doubled,” will change. Those who run stop signs probably don’t want the fine to double.
- (c) Answers vary.
11. Based on the information, scheme A will be better because the blocks are similar. The plots bordering the river should be similar, and the plots away from the river should be similar.

Chapter Review Problems

1. Using a random-number table to select numbers for a Sudoku puzzle would be very inefficient. It would be much better to look at existing numbers that meet the puzzle’s requirements and eliminate numbers that don’t work.
2. Alisha’s study has a few problems and results will be anecdotal. For instance, it’s not clear that the puzzles she wants to download are all of the same difficulty level. Her friends willing to participate will likely have different levels of experience with the puzzles. Her friends are also volunteers and the self-timing may lead to some inaccurate measurements.
3. (a) Stratified.
 (b) Students on your campus with work-study jobs.
 (c) Number of hours scheduled to work each week; Quantitative; Ratio.
 (d) Applicability to future employment goals, as measured by the scale given; Qualitative; Ordinal.
 (e) Statistic.
 (f) The nonresponse rate is 60%, and yes, this could introduce bias into the results. Answers vary.
 (g) No, since the students were only drawn from one campus, then the results of the study would only generalize to that campus, if the data were collected using randomization.
4. The implied population is all the listeners (or even all the voters). The variable is the voting preference of a caller. There is probably bias in the selection of the sample because those with the strongest opinions are most likely to call in.
5. Using the random-number table, pick seven digits at random. Digits 0, 1, and 2 can correspond to “Yes,” and digits 3, 4, 5, 6, 7, 8, and 9 can correspond to “No.” This will effectively simulate a random draw from a population with 30% TIVO owners.
6. (a) Cluster. (b) Convenience. (c) Systematic. (d) Simple random. (e) Stratified.
7. (a) This was an observational study because the researchers did not apply a treatment.
 (b) This was an experiment because the two groups were given different tests and the results were compared.
8. (a) Randomly select 500 donors to receive the literature and 500 donors to receive the phone call. After the donation collection period, compare the percentage who donated from each of the two treatment groups. A placebo is not being used.
 (b) Randomly select the 43 adults to be given the treatment gel and the 42 adults to receive the placebo gel. After the treatment period, compare the whiteness of the two groups. To make this double blind, neither the treatment administrators nor would the patients would know which gel the patients are receiving.
 (c) Before assigning donors to the literature or the phone call, first block them into the three age groups. In each age group, half would receive the literature and half would receive the phone call. Compare the amounts received within each block.

9. Answers vary. Questions should be worded in a clear, concise, and unbiased manner. No questions should be misleading. Commonsense rules should be stated for any numerical answers.
10. No response required.
11. (a) This is an experiment; the treatment was the amount of light given to the colonies.
(b) The control group is the colony exposed to normal light, while the treatment group is the exposed to continuous light.
(c) The number of fireflies living at the end of 72 hours.
(d) Ratio.

Chapter 2: Organizing Data

Section 2.1

1. Class limits are possible data values, and they specify the span of data values that fall within a class. Class boundaries are not possible data values; they are values halfway between the upper class limit of one class and the lower class limit of the next class.
2. Each data value must fall into one class. Data values above 50 do not have a class.
3. The classes overlap. A data value such as 20 falls into two classes.
4. These class widths are 11.
5. $\text{Width} = \frac{82 - 20}{7} \approx 8.86$, so round up to 9. The class limits are 20 – 28, 29 – 37, 38 – 46, 47 – 55, 56 – 64, 65 – 73, 74 – 82.
6. $\text{Width} = \frac{120 - 10}{5} = 22$, so round up to 23. The class limits are 10 – 32, 33 – 55, 56 – 78, 79 – 101, 102 – 124.
7.
 - (a) The distribution is most likely skewed right, with many short times and only a few long wait times.
 - (b) A bimodal distribution might exist if there are different wait times during busy versus slow periods. During the morning rush, many long wait times might occur, but during the slow afternoon, most wait times will be very short.
8. The data set consists of the numbers 1 up through 100, with each value occurring once. The histogram will be uniform.
9.
 - (a) Yes.
 - (b)

